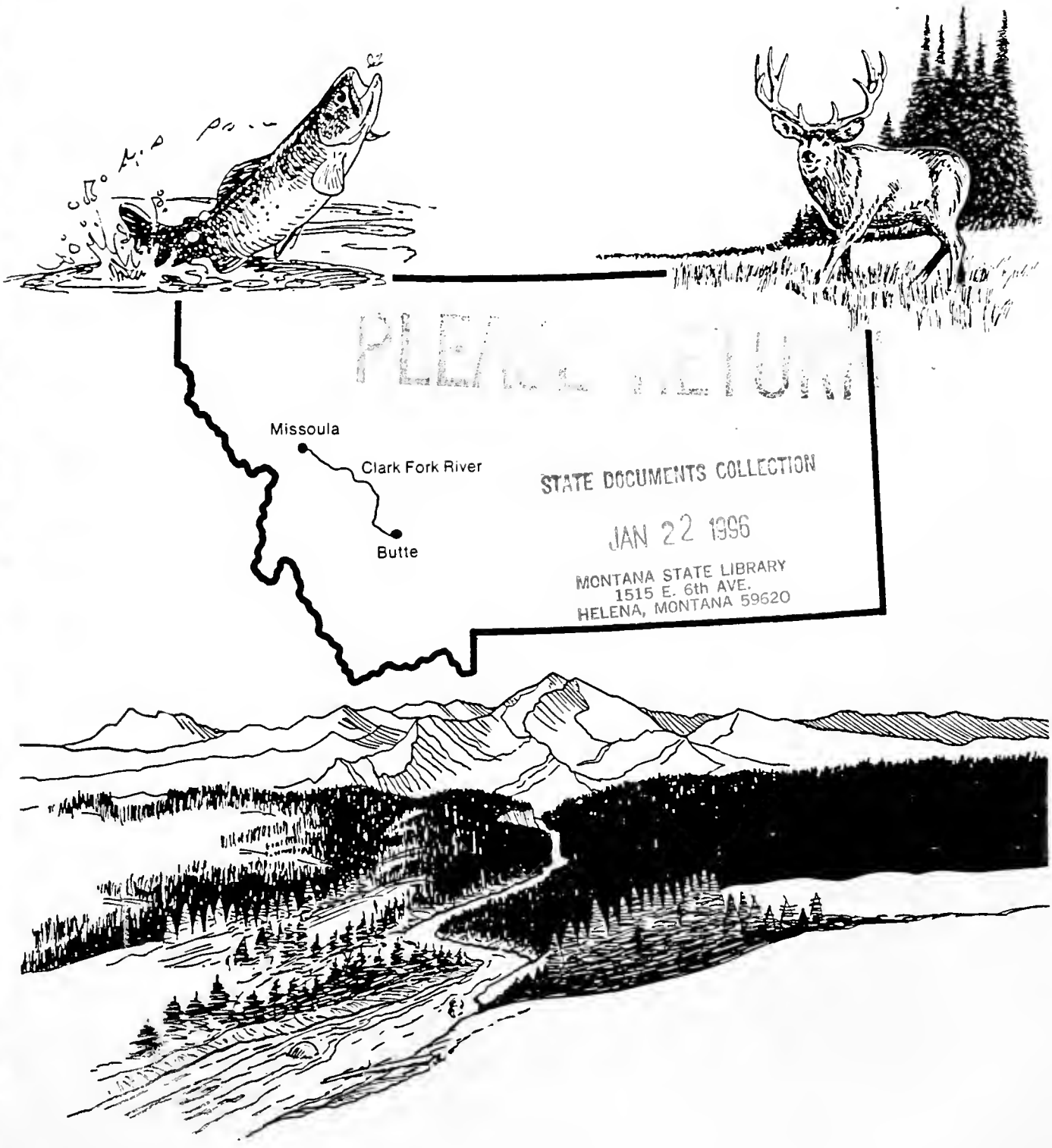
S Hagler Bailly
363.738 Consulting
H2rrcvn Rebuttal report
1995

TATE OF MONTANA
NATURAL RESOURCE DAMAGE PROGRAM

REBUTTAL REPORT:
CONTINGENT VALUATION OF NATURAL RESOURCE DAMAGES DUE TO INJURIES TO THE UPPER CLARK FORK RIVER BASIN

OCTOBER 1995



PLEASE RETURN
STATE DOCUMENTS COLLECTION
JAN 22 1996
MONTANA STATE LIBRARY
1515 E. 6th AVE.
HELENA, MONTANA 59620

MONTANA STATE LIBRARY
S 363.738 H2rrcvn 1995 c.1
Rebuttal report :contingent valuation of
3 0864 00096637 7

Water Mill,
363.7 Consulting
H2rrcvn Rebuttal report
1995

REBUTTAL REPORT:

CONTINGENT VALUATION OF NATURAL RESOURCE DAMAGES DUE TO INJURIES TO THE UPPER CLARK FORK RIVER BASIN

Prepared for:

State of Montana
Natural Resource Damage Program
1310 East Lockey Avenue
Helena, MT 59620

Prepared by:

Hagler Bailly Consulting, Inc.
P.O. Drawer O
Boulder, CO 80306-1906
(303) 449-5515

William D. Schulze
Robert D. Rowe
Gary H. McClelland
Donald Waldman

October 18, 1995

CONTENTS

Tables
Figures
Acronyms

Chapter 1 Introduction

Chapter 2 The Reliability of Contingent Valuation

2.1 The General Reliability of Contingent Valuation 2-1
2.2 Demand Curves and the Sensitivity of WTP to Scope 2-4
2.3 Is No Number Better than a CV Number? 2-8

Chapter 3 Plausibility of the Montana CVM Values

3.1 Values and Opinions of Montana Residents 3-1
3.2 Substitute Sites 3-2
3.3 Baseline Site Quality 3-3
3.4 Incompatibility of Stated Values and Cash Contributions 3-4

Appendix 3A Verbal Protocol and Pretest Comments on the Validity of WTP Responses

Chapter 4 Scope Testing and Embedding

4.1 The Scope Test and Embedding Issues 4-1
4.2 Analysis 4-4
4.3 Other Related Issues 4-9

Chapter 5 Scenario Definitions

5.1 Substitutes 5-1
5.2 Commodity Definition 5-2

Chapter 6 WTP Elicitation

6.1 Ten-Year Payment Period 6-1

6.2 Payment Vehicles 6-3
6.2.1 The Payment Question 6-3
6.2.2 The Payment Vehicles 6-3
6.2.3 Payment Card Issues 6-7
6.3 Incentive Bias 6-9

Chapter 7 Sampling and Statistical Issues

7.1 Nonresponse Bias Correction 7-1
7.2 The State's WTP Regression Analysis 7-2
7.2.1 Tobit Model Specification 7-2
7.2.2 Indices 7-2
7.3 Data Treatment Issues 7-4
7.3.1 Missing Data 7-5
7.3.2 Coding of Second WTP Responses 7-5
7.3.3 Protest Zeros and Large WTP Responses 7-6
7.4 Responsibility Bias Correction 7-7

Appendix 7A Sample Pretest Comments Concerning Partial Versus Complete Cleanup

Chapter 8 Other Contingent Valuation Issues

8.1 Conservative Estimates 8-1
8.2 Aggregate Damage Computations 8-1

Chapter 9 References

TABLES

4-1 Between-Sample and Within-Sample Comparisons of the Difference in WTP for Complete and Partial Cleanup of the Upper Clark Fork NPL Sites 4-7

6-1 Comparisons of Values Measured Using WTP Questions with Dichotomous Choice Questions 6-4

FIGURES

2-1 The Individual Demand Curve 2-5
2-2 Demand Curve for Seals 2-7
2-3 Demand for Cleanup of the Upper Clark Fork River Basin 2-9

ACRONYMS

CV	contingent valuation
CVM	contingent valuation method
DOI	Department of the Interior
EEAC	Environmental Economics Advisory Committee
EPA	Environmental Protection Agency
NPL	National Priorities List
NOAA	National Oceanic and Atmospheric Administration
WTP	willingness to pay

CHAPTER 1
INTRODUCTION

In this report the State of Montana provides a rebuttal to ARCO's Expert Reports by Jerry Hausman (1995) and William Desvousges (1995) critiquing the State of Montana's contingent valuation study (Schulze et al., 1995).[1] This rebuttal makes four major points, as summarized below.

First, the opinions of ARCO's experts are not based on the refereed scientific literature on contingent valuation (CV). In particular, Dr. Hausman's characterization of CV as unreliable is in direct contradiction to overwhelming scientific evidence published in economics journals. Further, his views on CV have never been published, to our knowledge, in a refereed economics journal and so, in contrast to the views of the State's experts, have never been subjected to peer review. ARCO's experts misrepresent both the weight of the evidence in the scientific literature as well as the conclusions of the NOAA Panel on CV. Often, the comments by ARCO experts are simply speculative possibilities with no substance or supporting data.

Second, ARCO's experts use misleading arguments with respect to the Montana CV study. For example, Dr. Hausman's key assumption about scope testing is incorrectly applied to the Montana CV study; and his argument that the value of additional cleanup should increase proportionately, more than the amount of cleanup increases violates the economic "law of diminishing marginal returns," which is that as the level of cleanup increases, people will pay less for further increments in cleanup.

Third, in areas where scientific debate exists regarding CV methodology, the Montana contingent valuation study uses highly conservative approaches. In several cases, ARCO's experts attack the use of what the scientific literature shows to be conservative approaches.

[1] Incomplete and late document production by ARCO has prevented a full understanding of ARCO's experts' opinions and has hampered preparation of this rebuttal report. Specifically, materials generated, reviewed, or considered by Dr. Jerry Hausman were only partially provided to the State of Montana prior to the final preparation of this report, which precluded full review for the rebuttal report.

Finally, ARCO's experts, in essence, argue that the citizens of Montana are incapable of determining their own values for cleaning up the upper Clark Fork NPL sites.[2] Rather, they argue that they should make that determination for the citizens of Montana, and their determination is that the value to Montana citizens of cleaning up the upper Clark Fork sites should be very small, if not zero.

[2] A footnote on page 12 of Dr. Hausman's report implies that he and like-minded economists do not believe that the citizens of Montana are capable of even voting in a manner consistent with their economic preferences. It appears that Dr. Hausman would find invalid any means by which the citizens of Montana might try to express their values. In contrast, the State believes its citizens are perfectly capable of voting their interests and that their internally consistent responses with corresponding consistent written comments indicates that they were capable of expressing their preferences in the CV survey.

CHAPTER 2
THE RELIABILITY OF CONTINGENT VALUATION

2.1 THE GENERAL RELIABILITY OF CONTINGENT VALUATION

Dr. Hausman begins his Expert Report with a critical appraisal of the Contingent Valuation (CV) method in general (pages 4 through 23). He follows his general views with specific comments on the Montana CV study (Dr. Desvousges follows the same organization and makes many of the same general comments on CV as Dr. Hausman. Therefore, we focus on Dr. Hausman's comments here). The general views expressed by Dr. Hausman of the CV method are largely unsubstantiated and rely almost exclusively on an article published in the *Journal of Economic Perspectives* (1994) that he co-authored with P. Diamond. The *Journal of Economic Perspectives* only publishes invited papers on timely topics without the usual peer-review process. Thus, this article as well as the only other two publications on CV that he has co-authored (Hausman and Diamond, 1993 and Hausman et al., 1993) were not submitted for the scrutiny of peer-reviewers. Dr. Hausman's personal opinions are for the most part in direct contradiction with the substantial peer-reviewed economic literature on CV. They do not represent the majority of published literature on CV, they do not reflect the opinions of the NOAA Blue Ribbon Panel on the use of CV for Damage Assessments (Arrow et al., 1993), and they do not correspond to the substance of several administrative and legal pronouncements on the validity of CV (see the State's CV report, Section 1.3).

In Section III of his report, Dr. Hausman makes several assertions that CV is an unreliable method. He repeatedly writes that CV does not measure economic preferences and that it is not a valid method for measuring use or nonuse values (pages 4, 5, and 8). He asserts that CV is inferior to the methods that rely on observed behavior that economists generally use because its results cannot be tested against actual consumer behavior (page 6). In fact, he goes further by stating that no tests of CV's validity exist (page 7).

Dr. Hausman is factually wrong on all counts. The peer-reviewed economics literature contains numerous studies comparing CV to standard economic methods that use real market behavior to infer values.[1] The first thorough review of comparative studies was prepared by Cummings, Brookshire, and Schulze (1986). It formed the starting point for a conference on assessing CV and was later published in a book. In their conclusion, the authors note:

[1] See for instance, Knetsch and Davis (1966); Desvousges, Smith, and McGivney (1983); Brookshire, Thayer, Schulze, and d'Arge (1982); Seller, Stoll, and Chavas (1985); Cummings, Brookshire, and Schulze (1986); Bishop and Heberlein (1990); Dickie, Fisher, and Gerking (1987); Brookshire and Coursey (1987); and Thayer (1981).

> "our final assessment of the state of the arts of the CVM is generally positive. We find impressive the accuracy of CVM measures inferred by the available evidence at this stage of the method's development" (page 234).

Cummings, Brookshire, and Schulze identified approaches for validating CV by using direct and indirect market-based methods and by using laboratory methods. A brief summary of studies using these approaches will give a flavor for what the literature actually indicates about the validity of CV.

Validation Using Travel Costs

Knetsch and Davis, in a study published in 1966, describe what is generally regarded as the first CV study. They examined the value of recreation and, since the traditional method for valuing recreation uses travel costs (which are obtained from actual market behavior) to estimate values, they also compared their CV results to those obtained from a travel cost study. Both approaches gave similar value estimates. Over the nearly thirty years since this first study, many additional studies have compared travel costs with CV with generally favorable results (see, for example, Desvousges et al., 1983; Adamowicz and Graham-Tomasi, 1991; Loomis et al., 1986; Seller et al., 1985; and Walsh et al., 1990).

Validation Using Field Experiments

Bishop and Heberlein (1979), in a field experiment, compared values for goose hunting permits obtained through CV with the behavior of individuals who were offered actual monetary payment by the experimenters to return their permits. Again comparable estimates of value were obtained. It should be noted that, although this study is often cited, it was published in a proceedings issue of the *American Journal of Agricultural Economics* and was not given the usual peer review. However, the study's prominence in the literature suggests a high regard for the work among peers (the study has been cited more than one hundred times in the refereed literature). A similar experiment was published by Dickie et al. (1987) in the *Journal of the American Statistical Association* showing that hypothetical values for purchasing fresh strawberries were similar to actual values.

Validation Using Property Values

Another method for validating CV is to compare values obtained using CV to those revealed by the hedonic property value market. Brookshire, Thayer, Schulze, and d'Arge published the first study of this type in the *American Economic Review* in 1982. They showed that the additional cost of buying a home in an area with cleaner air corresponded with what people said they would pay for cleaner air in a CV survey. This general approach has been used many times, again with positive results (Brookshire, Thayer, Tschirhart, and Schulze, 1985; d'Arge and Shogren, 1989; and Blomquist, 1988).

Validation Using Laboratory Experiments

Experimental economists can create real markets in a laboratory environment where subjects can buy and sell commodities under carefully controlled incentive compatible conditions so that true values are revealed. These true values can then be compared to the results of a CV survey asking subjects to value the same commodity. The first study to use this approach was published by Coursey, Hovis, and Schulze, in the *Quarterly Journal of Economics* in 1987. They used as a commodity the tasting of a safe but bitter liquid (sucrose octa acetate) commonly used by psychologists in taste experiments. Contingent values to avoid the taste experience were in close agreement with actual money values. A substantial number of similar experiments have been published since that time (see, for instance, Irwin, McClelland, and Schulze, 1992; McClelland, Schulze, and Coursey, 1993; Boyce et al., 1992; Shogren et al., 1994) and are summarized in Schulze et al. (1995) and show that CV provides comparable values when normal statistical procedures are used to eliminate outliers, such as the procedures used in the State's CV study.

Three Recent Reviews of CV

Three recent independent literature reviews of studies comparing CV to market methods (also referred to as observed behavioral and indirect methods) are available (Smith, 1993; Bishop and Heberlein, 1990; and Carson et al., 1994). All reach the same conclusion: that CV provides measures of value that are comparable to market methods

In the first review, Bishop and Heberlein (1990) state: "The results of such comparisons have been positive. For instance, contingent values have been shown to be roughly comparable to values derived from travel-cost models. . ., hedonic price models. . ., and the costs and prices of substitutes" (page 93).

The review by Smith (1993) also included several studies that compared the results of hypothetical CVM studies with the actual behavior of consumers in field and laboratory experiments. He concludes that "these studies support the CVM estimates as being comparable in performance to the alternative approach providing the reference point" (page 9).

In the most complete evaluation yet, Carson et al. (1994) reviewed over 80 comparative studies containing in excess of 600 comparisons between contingent valuation and indirect methods (e.g., travel costs, property values, and market studies). Hanemann (1994) cites this review and concludes that "the results are often fairly close". Overall, the contingent valuation estimates are slightly *lower* than the revealed preference estimates and highly correlated with them" (pages 29-30).

It is possible to argue that the majority of these comparison studies are related to use values rather than nonuse values. A major difference between use and nonuse values is familiarity

with the commodity. Valuation is more difficult when one is less familiar with a commodity, and easier when one is very familiar with a commodity (e.g., anglers valuing changes in fishing quality)[2]. In this context, an important question is how familiar are Montana residents with the commodity we are asking them to value, including both use and nonuse values? These commodities, usable groundwater, an unpolluted stream, and clean land, are commodities with which most Montana residents are familiar and use (including fishing, hiking, and active and casual sightseeing). Also, the CV survey indicates that a substantial majority (89 percent) of Montana residents are familiar with one or more of the specific upper Clark Fork NPL sites (CV report Table 4-1).[3] Accordingly, residents of Montana are familiar with the commodity presented and can be expected to express reliable value for these natural resource injuries.

As this evidence shows, Dr. Hausman is factually incorrect in several of his statements. Contrary to what he writes:

- Validity tests of CV exist and many such studies have been published in peer-reviewed publications;
- It has been demonstrated that CV is capable of measuring preferences and values, and that CV performs as well as traditional methods used by economists;
- CV tends to provide lower estimates of value than the alternative methods generally used by economists, contrary to the claim by Hausman (page 5) that CV is upwardly biased;
- As demonstrated in the literature, the hypothetical nature of CV surveys does not prevent CV from providing accurate measures of preferences and value or appropriately predicting behavior (see next section).

2.2 DEMAND CURVES AND THE SENSITIVITY OF WTP TO SCOPE

In his appraisal of the CV method in general and the State's application of the CV method, Dr. Hausman discusses the issue of how WTP should vary with changes in the scope of the commodity being valued (later he discusses scope issues with specific reference to the Montana CV study, see Section 4 below).

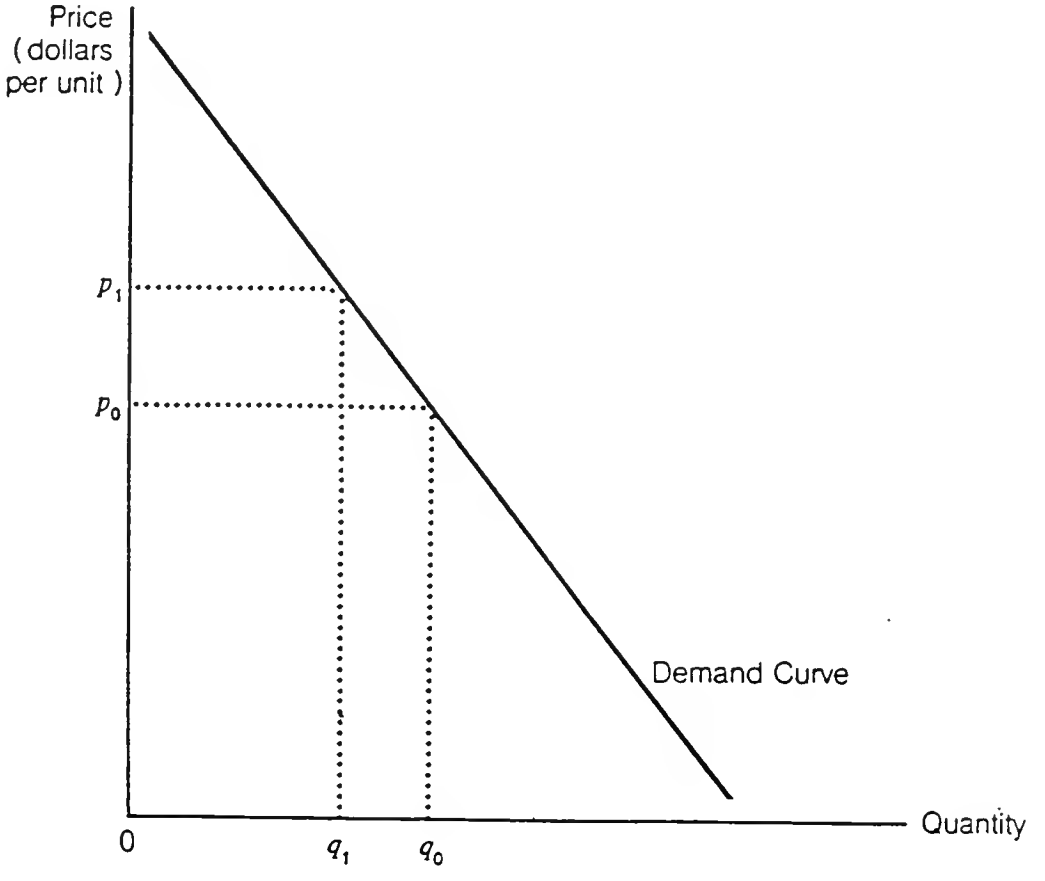
---

[2] Dr. Desvousges also notes this (page 16) where he indicates that the margin of error is greater in nonuse value studies than use value studies.

[3] Respondents who were not familiar with these sites provided WTP responses that were half as large as the WTP responses provided by other respondents, again indicating a consistent response pattern in the CV results (see also Section B.6 in the State's revised Determination of Damages report).

Dr. Hausman introduces the example of two programs designed to save seals from a sure death in order to argue that CV produces theoretically inconsistent results (pages 9-10). In a clever manipulation of economic theory, Dr. Hausman makes the suggestion that a program to save 1,000 seals from death should be valued at least ten times more than a program that would save only 100 seals. Dr. Hausman's example is irrelevant in assessing the State of Montana CV study and is misleading. The example is not pertinent because it discusses the value of avoiding losses in environmental quality, as opposed to the value of improving (or adding to) environmental quality as is done in the State of Montana CV study. The Montana CV study evaluates the benefits associated with partial and complete cleanup of the upper Clark Fork Basin by focusing respondents on environmental quality improvements that each

Figure 2-1
The Individual Demand Curve



program would bring. Using the seals example introduced by Dr. Hausman, this is akin to measuring the value of two different programs seeking to increase the seal population by 100 and 1000 animals respectively.

To see why the difference between avoiding losses and improving environmental quality matters, a look at the economic theory of demand is helpful. One of the fundamental verified predictions of economic theory is that the number of units of a good that a consumer wants to buy decreases as the price of the good goes up. This assumption gives rise to a downward sloping demand curve like the one depicted in Figure 2-1 (reproduced from Tietenberg, 1994). This graph shows that the maximum amount that a consumer is willing to pay for the first unit of a good is higher than for the second unit, and that the value decreases for each additional unit.

In this context, it is easy to show that the value of increasing the population by 1,000 seals should be strictly less than 10 times the value of the adding 100 seals. Suppose that the current seal population stands at 10,000 animals. Then, according to Figure 2-2[4] increasing the population by 100 seals has a total value of $9.50 (the area under the curve between 10,000 and 10,100). Since successive increments have a decreasing value, another program that would add another 100 seals to increase the population from 10,100 to 10,200 will be valued at less than $9.50. According to Figure 2-3, the second group of 100 seals is valued at $8.50. Thus, a program that would add 200 seals would be worth $9.50 + $8.50 = $18 to the consumer. This is less than twice the value of the initial proposal to increase the population by 100 seals.

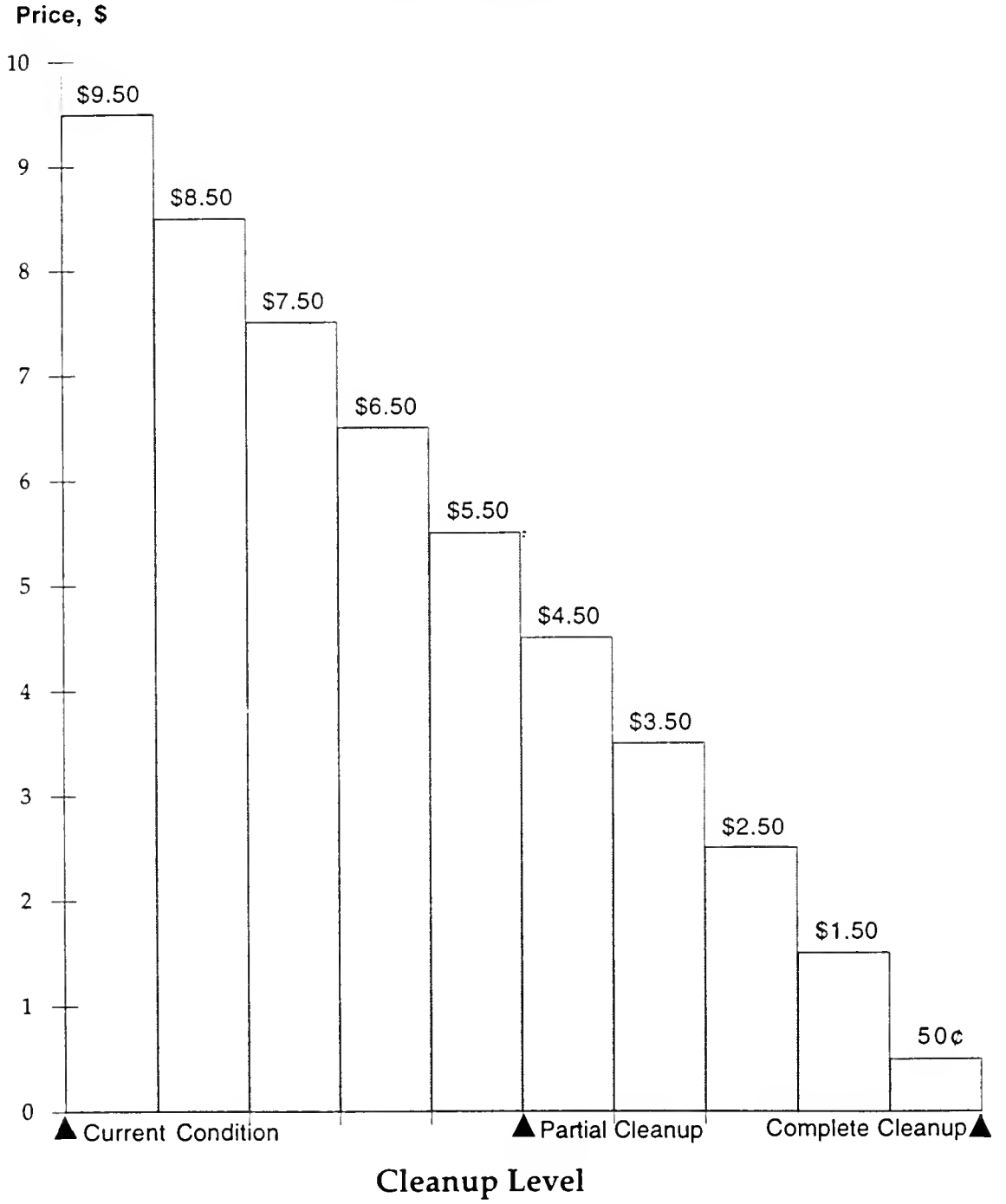
Continuing this reasoning, we find that increasing the population from 10,900 to 11,000 is valued at $0.50. In this example, the total value of increasing the seal population by 1,000 would be $50.00, which is less than six times (not more than ten times, as Dr. Hausman suggests) the value of the initial increase of 100 animals.

Because of the standard assumption of downward sloping demand curves, economic theory predicts unambiguously that a program of environmental improvement ten times larger than another similar program will have a value less than ten times the value of the smaller program. Further, theory does not put a bound on how close the values for a small and a large program should be.

This is illustrated in Figure 2-3, in which the same demand curve (as used for seals) is applied to the Montana CV study. In the graph, current conditions at the upper Clark Fork Basin are analogous to the case of 10,000 seals, and complete cleanup is analogous to the case of 11,000 seals. Partial cleanup of the upper Clark Fork Basin corresponds on this graph to 10,500 seals. As calculated for the seal case, the value corresponding to complete cleanup

[4] Figure 2-2 is a reproduction of Tietenberg's Figure 2.3 which has been relabelled for this example.

Figure 2-2
Demand Curve for Seals



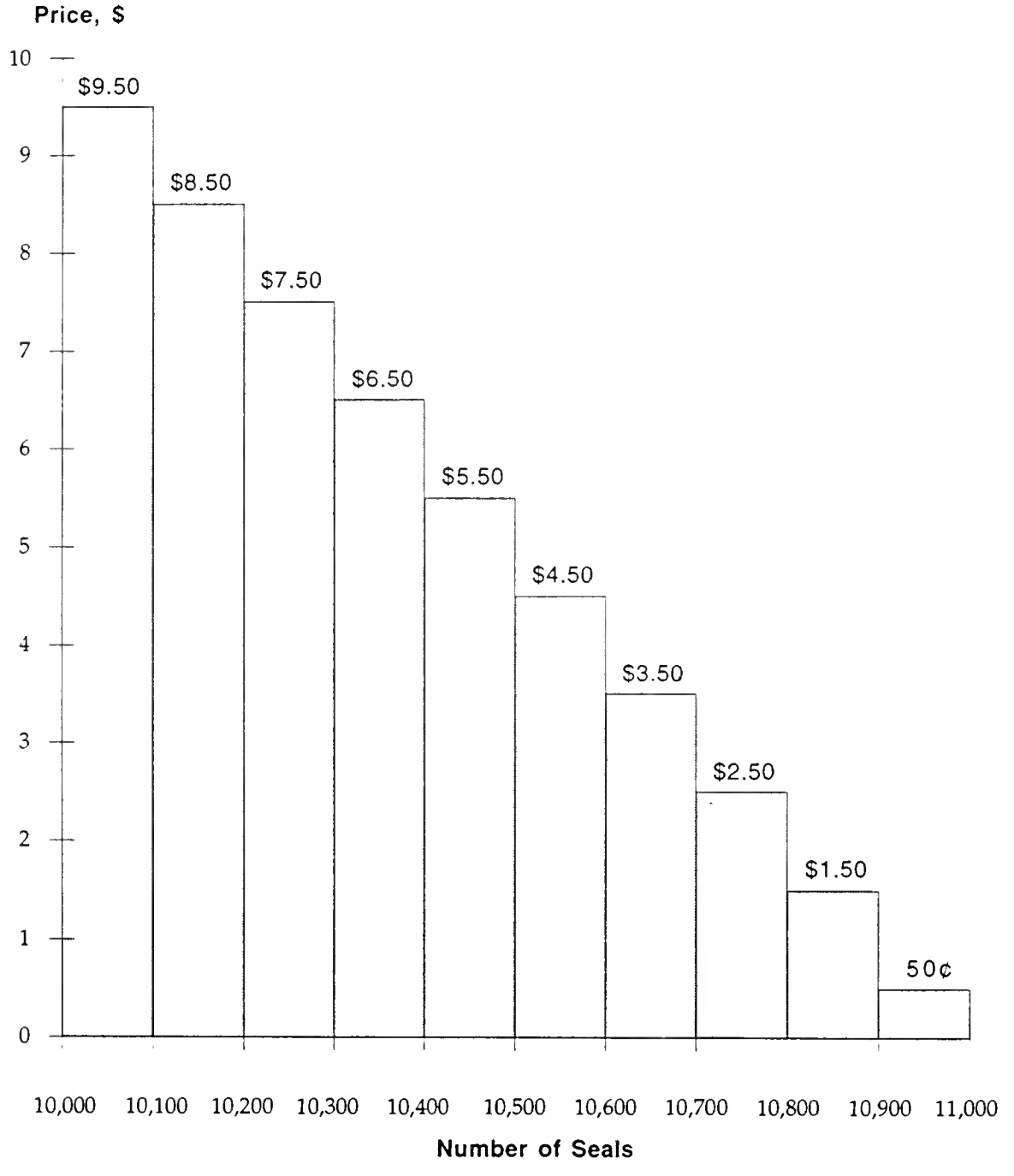
is $50 (similar to the $49 value in the Montana CV study). The value for partial cleanup would be $37.50 (again, similar to the $36 partial cleanup value in the Montana CV study). The difference between complete and partial cleanup is equivalent to the residual damages of $12.50. This corresponds to the structure of the Montana CV study, and demonstrates that the Montana CV results are completely consistent with a downward sloping demand curve as predicted by classical economic theory, and are not irrational or unusual as Dr. Hausman claims.

In conclusion, Dr. Hausman's argument concerning scope is misleading in that it compares programs that prevent the loss of environmental quality rather than obtaining environmental improvements. The difference in values for different scopes of the program is not only determined by the objective differences in the size of two programs (for example the exact number of seals to be added to the population) but also by the extent to which individuals perceive the programs as being different. Individuals may not perceive an appreciable difference between saving 100 or 1,000 seals if 10 million seals already exist. In other words, if respondents to CV surveys do not regard two programs of different sizes as truly providing them with a different level of benefits, they should not be expected to value these programs differently. On this basis, Smith (1993) has criticized Diamond et al. (1992) and Desvousges et al. (1992) for hastily concluding (after they had observed some insensitivity to scope in their CV studies) that CV does not measure economic preferences or that it does not provide either valid or reliable estimates of nonuse damages. Such conclusions, he wrote, are not warranted.

2.3 IS NO NUMBER BETTER THAN A CV NUMBER?

Dr. Hausman writes that "with regard to compensatory damage estimates, some number is not better than no number when the number that exists is badly biased" (page 8). This statement is inaccurate and misleading. The available empirical evidence does not support the view that CV is badly biased. Suggesting that "no number" be used is a de facto proposal to value damages at $0.00. In the presence of substantial damages to public natural resources and the significant concern expressed by Montana residents about the issue (see the State's CV report) we can be confident that $0.00, rather than the State's CV estimate of damages, is the badly biased estimate of value.

Figure 2-3
Demand for Cleanup of the Upper Clark Fork River Basin



CHAPTER 3
PLAUSIBILITY OF THE MONTANA CVM VALUES

Dr. Hausman begins his critique of the specific Montana CV study with his opinion that the estimated average WTP of Montana households ($49 per year for 10 years) to clean up the upper Clark Fork River sites is "economically implausible" (pages 24-26). His opinions are speculative, are in direct contrast to the written opinions and values of the residents of Montana (Section 3.1), and are based on assessments of the site and injuries at the site that contradict the facts (Sections 3.2 through 3.4).

3.1 VALUES AND OPINIONS OF MONTANA RESIDENTS

An average WTP per Montana household of $49 per year for 10 years to clean up the upper Clark Fork NPL sites can hardly be considered "economically implausible." One should first remember that these individuals are valuing the cleanup, in their home state, of the largest Superfund complex in the country. A WTP of $49 per year is reasonable when one considers that this amounts to less than one-fifth of 1 percent of the average annual Montana household income.

In the CV survey, respondents often reaffirmed the reasonableness of their stated WTP values in their own written comments on the survey instruments. As stated on page 5-18 of the CV report, "The most prevalent type of comment (258 made) was simply an affirmation of the WTP response. These comments frequently referenced the respondent's income level as a determinant of WTP" (a sample of the verbal protocol and pretest comments on this issue is included in Appendix 3A).

The values that Montana residents place on cleaning up the upper Clark Fork NPL sites are consistent with their attitudes and behaviors. For example, residents rated cleaning up hazardous waste sites in Montana as a very important activity (CV report Table 4-2), and the upper Clark Fork sites were rated as the most important hazardous waste sites in Montana to clean up (Table 4-4). Cleaning up these sites will lead to enhanced recreation opportunities for over 70,000 resident anglers (and over 65,000 non-resident anglers) who are active in river and stream fishing in southwestern Montana (Morey et al., 1995), and also will provide nonfishing recreation benefits and groundwater benefits. The ARCO survey of Montana households also demonstrates the importance of maintaining and improving natural resources and recreation opportunities in Montana. Of all households, 91 percent participated in outdoor recreation in the past 12 months (Desvousges and Waters Appendix A to Volume III, page A-19, Table A-7 footnote). Of the Montana recreators, 88 percent agreed or strongly agreed with

the following statement "Outdoor recreation is an important part of my life." When asked about whether governments are spending the right amount, too much, or not enough on *improving* natural resources for outdoor recreation in Montana, recreators overwhelmingly indicated the same or more should be spent.[1] Clearly, even ARCO's survey of Montana residents indicates that Montana residents are concerned about improving natural resources. Finally, the literature is replete with studies demonstrating significant property value impacts (much larger than $49 per year) resulting from proximity to hazardous waste sites and landfills (e.g., Thayer et al., 1993; McConnell, 1993; Mendelsohn et al., 1992, Schulze, et al., 1986; and Smith and Desvousges, 1986[2]).

3.2 SUBSTITUTE SITES

Dr. Hausman asserts that the Montana residents should have small values for cleaning up the upper Clark Fork NPL sites because, he asserts, the Clark Fork sites are not "unique" and "many uninjured, relatively pristine rivers exist in Montana that would appear to be good substitutes for a cleaned up Silver Bow Creek and Clark Fork River" (page 24).

Dr. Hausman glosses over the fact that the effect of uniqueness on value is a matter of degree. First, a natural resource does not have to be the only one of its kind to have a value to some or many individuals. Second, the magnitude of resource impacts at the upper Clark Fork sites is unique. There are 140 miles of impacted river, which in baseline condition would support significant, highly valued recreational activities; and the terrestrial and groundwater damages at the upper Clark Fork sites are unique in that they are among the geographically largest areas of natural resource injury in the country. These natural resources, and the magnitude of the injuries, are sufficiently unique and important to Montana residents

---

[1] This question was asked separately for state and local government spending, and for federal government spending. We report the results adjusted for don't knows and refusals. For state and local government spending, 48% said spending was about right, 35% said not enough was being spent, and only 8% said too much was being spent. For federal spending, 33% said spending was about right, 48% said more should be spent, and only 9% said too much was being spent. (Desvousges and Waters Appendix A to Volume III, Table A-9).

[2] Smith and Desvousges (1986) conducted a contingent valuation survey of residents of eastern Massachusetts to quantify WTP to avoid living near a hazardous waste site. The average Massachusetts' homeowner was willing to pay *annually* $330-$495 ($1984) per mile to avoid living within a mile of a hazardous waste site. This is a capitalized asset value, based on a 13 percent interest rate, of about $2,470-$3,200 per house (in 1984 dollars) to relocate one mile away from the site. One can expect that property value impacts of the Clark Fork sites extend much farther than one mile from the sites. This is because the upper Clark Fork sites are geographically the largest sites in the country, because the sites can seen from great distances, and because the injured resources are important features of the region, unlike a hazardous waste site in eastern Massachusetts.

that they place a value on returning them to baseline condition, even if Dr. Hausman does not.

Dr. Hausman is correct that increasing substitutes may reduce value. For example, as distance from the site increases, the number of substitutes increases, and we expect WTP for cleanup of the upper Clark Fork sites to decrease, but it need not be $0. This economic tenet is reflected in the Montana CV results where the average household WTP decreases from $57 per year for households within 100 miles of the site to $18 per year for Montana households farther than 500 miles from the site (CV Report,Table 5-14). This reflects that uniqueness is a matter of degree, and that the CV results are consistent with economic expectations.

3.3 BASELINE SITE QUALITY

Dr. Hausman asserts that even if the upper Clark Fork NPL sites were cleaned up, the improvement would not be large and the site quality would be low (page 24). For example, he states that there would still be a highway and a railroad in the river corridor, there would still be other discharges to water bodies, and Silver Bow Creek would not be a pristine wilderness creek. These comments contradict and misrepresent the facts of the case. For example, if the sites are returned to baseline conditions:

- About 15 square miles of terrestrial acreage, now with limited or no vegetation, would have normal vegetation and habitat.
- Hundreds of thousands of acre-feet of groundwater would no longer be contaminated.
- The 27-mile-long Silver Bow Creek, which has limited vegetation over about 750 acres and which does not support aquatic life, would have normal vegetation, and would have fish stocks per hectare that would exceed current stocks in all but a few miles of the upper Clark Fork river between Warm Springs Ponds and Milltown (State's aquatic injury assessment).
- Fish stocks in about 115 miles of the upper Clark Fork River would increase nearly five-fold, which would place these sites among the better coldwater recreational fishing sites in southwestern Montana. Riparian habitat would be improved for about 250 acres along the upper Clark Fork River. The recreation analysis indicates that these improvements would result in about a doubling of recreation use at these sites.
- Roads and/or railroads occur near most major rivers in Montana and these rivers receive considerable recreation use, in part because the roads provide access and because anglers are more concerned with the type, number, and size of fish they catch (Morey et al., 1995 Table 3-1). For example, Highway I-15 runs parallel to portions of the Big Hole and Missouri rivers, I-90 runs parallel to the Galatin and Yellowstone rivers, and state roads parallel virtually every other major river.

3.4 INCOMPATIBILITY OF STATED VALUES AND CASH CONTRIBUTIONS

Dr. Hausman questions the validity of the stated WTP by Montana residents because of evidence he cites that stated values may exceed actual cash contributions (page 25). These comments are not tied to the case and lack scientific foundation.

First, Dr. Hausman notes that only 0.5% of Montana 1993 tax returns made an actual payment to the nongame wildlife program, which resulted in an average payment per household of $0.05, almost 1,000 times smaller than the Montana CV mean WTP for cleaning up the upper Clark Fork sites. From this, he infers that WTP for the cleanup of the upper Clark Fork sites is seriously overstated, even through the two programs are not related.

The tax contribution data simply indicate the serious limitations in using voluntary cash contributions as a basis to determine values. In fact, economic evidence suggests that such mechanisms encourage free riding — a failure to contribute based on the hope that others will pay enough to provide the public good (see Hagel and Roth, 1995; Isaac et al., 1985; and Kim and Walker, 1984). This free riding phenomenon implies that such contribution mechanisms will invariably undervalue the public good. In fact, the tax data suggest that fewer than 2,000 Montana households value protecting nongame wildlife. This is in sharp contrast with recreation statistics for the state. For example, the 1985 National Survey of Fishing, Hunting, and Wildlife-Associated Activities estimated that 241,000 people observed wildlife, 219,400 people fed wildlife, and 64,100 people took pictures of wildlife in 1985 (see also other recreation data in Chapter 8 of Morey et al., 1995 and the discussion in Section 3.1 of the ARCO survey findings on Montana residents about the importance of improving natural resources for outdoor recreation in Montana). Clearly, it is the tax contribution data that provide an unreliable measure of value and these data should not be relied on in any way to infer values for nongame wildlife protection, and certainly not relied on to make inferences about the value of cleaning up the upper Clark Fork NPL sites, or the reliability of the State's CV study.

Second, Dr. Hausman cites two studies (page 25) to suggest that CV WTP values exceed cash payments in "paired comparisons" (Duffield and Patterson, 1991, 1992; and Seip and Strand, 1992). These studies use methods that provide a poor basis for evaluating contingent valuation (Hanemann, 1994). Both studies solicit contributions (and thus are not incentive compatible and invite free-riding in the cash payments versions) rather than mimic a true CV format, and both studies have unmatched survey implementation procedures across the cash payment and CV treatments. Yet, in Duffield and Patterson the WTP results are comparable for those individuals who responded to the cash treatment and to the contribution treatment (the difference arises because of different response rates).

APPENDIX 3A
VERBAL PROTOCOL AND PRETEST COMMENTS ON THE VALIDITY OF WTP RESPONSES

"My household includes just me. I would be willing to pay $300.00 total a year for 10 years." Raw WTP=$300. 11/05/92-V1, S4, 005

"Currently our household income is small. It should increase over the next 10 years (after we graduate university). As our income increases, so would the amount we would be willing to pay." Raw WTP=$90. 11/05/92-V1, S4, 013

"That would be $400.00 over the next ten years which is acceptable." Raw WTP for complete cleanup is $40 per year. 11/05/92-V1, S3, 023

"I am no doctor or lawyer. So I am not a rich dude." Raw WTP=$125. 11/05/92-V1, S3, 022

"I think that $450 is a reasonable amount for a family to pay..." Raw WTP=$450. 11/05/92-V1, S3, 008.

"I feel that is about all I can afford on my income." Raw WTP=$125. 11/05/92-V1, S3, 006.

"My grandchildren must have the opportunity to live in a clean, safe environment and they shouldn't have to pay for the messes we allowed to happen. The $650 figure is less than we pay in property taxes, so that seemed reasonable..." Raw WTP=$650. 11/05/92-V1, S2, 006.

"I live in Helena, the economy is not high, does not fluctuate much, I can't foresee [sic] any significant increase in my income to warrant more fees-taxes than indicated." Raw WTP=$90. 11/05/92-V1, S1, 017.

"Because of the size of Household income I can't afford a large amount." Raw WTP=$40-$90 (three values circled). 11/05/92-V1, S1, 006.

"I think it is reasonable to say that most households, including mine could afford $125.00 a year for clean-up that would benefit all of us and our children and future grandchildren. It is everyone's responsibility." Raw WTP=$125. 11/05/92-V1, S1, 003.

"First, I am poor..." Raw WTP=$10. 11/05/92-V2, S4, 002.

"It is such a major problem it should be addressed and dealt with - If my contribution would be well utilized I say 'go for it'." Raw WTP=Don't Know. 11/05/92, S1, 009.

"I feel on my retirement income all I can afford is $450. I would like to see the Clark Fork NPL looking and being better." Raw WTP=$450. 11/05/92-V2, S1, 012.

"We have very little disposable income and would be hesitant to spend it on problems that do not impact us personally." Raw WTP=$25. 11/05/92-V2, S3, 007.

"I don't have a very high income (retired teacher of 17 years), but think the environment is exceedingly important and worth sacrificing for." Raw WTP=$450. 11/05/92-V2, S3, 010.

"I am a student with a limited income, yet I would donate time and money to help clean the Clark Fork and other areas, and believe all Montanans should." Raw WTP=$300. 11/05/92-V2, 11/05/92-V2, S3, 014.

"As a student I would be willing to pay less than if I was employed full time"; $90 WTP; 12/14/92 - V3; 6

"This was based on my income - I feel others who are in a higher income bracket could well afford to pay more for such an important project!!"; $40 WTP; 12/14/92 - V3 5

"Amount paid should not be a financial burden on most households."; $60 WTP; 12/14/92 - V3; 5

"This is a great concern for my family. We have traveled the US and Montana is our home. We have returned here to start a family and be in the surroundings that we love. I can't stand the fact that our home and all the wildlife that share this land with us is deteriorating. This is a very talked about subject in our family and believe any efforts towards clean-up are extremely important and needed."; $2,250 WTP; 12/14/92 - V3; 5

"I am currently on AFDC and going to school our money is short but I am willing to volunteer what skills I have."; $40 WTP; 12/14/92 - V3; 5

"As a household I am only accounting for myself because I am a student living with 3 other people my own age and I don't know how much they would contribute. I would increase that amount after I graduate from college to at least $200 which would be next year."; $125 WTP; 12/14/92 - V3; 4

"This is necessary for our children!"; $125 WTP; 12/15/92 - V2; 1

"$125 is about 10% of my home owner property tax. It seems a reasonable share."; $125 WTP; 12/15/92 - V2; 2

"With my income that I currently bring in I could handle about $200 a year"; $200 WTP; 12/15/92 - V2; 4

"A lot would depend on my financial status is related to job and family status. I would hope that I could pay more than $25 - but that is the most I could commit to for 10 years not knowing my future circumstances"; $25 WTP; 12/15/92 - V2; 3

"We will all have to help pay to keep the environment clean"; $200 WTP; 12/06/92 - V1; 1

"My family believes in commitment to the environment"; $200 WTP; 12/15/92 - V2; 3

"Any cleanup is better than none, $10-20 per month wouldn't be too bad for a better environment"; $200 WTP; 12/15/92 - V2; 3

"Although I don't have much money to spare, the Clark Fork gives me food and enjoyment from fishing and hunting throughout the year."; $125 WTP; 12/06/92 - V1; 6

"Currently I can not afford more I would pay more if my income would allow it";$200 WTP; 12/06/92 - V1; 5

"I am single - I live on a very tight budget with all prices increasing - I fish, hunt, hike, & love MT outdoors - $200 would be a little over what I could afford but I would be willing to make other cuts to see the complete cleanup take place"; $200 WTP; 12/06/92 - V1; 2

"[I would be willing to pay the amount of my property tax]; $450 WTP; 12/06/92 - V1; 1

"Soon, were (sic) going to pay $12/year for water quality disturb in Missoula & $10/year more seems fair"; $10 WTP; 12/06/92 - V1; 4

"I am a single household - earning minimum wage - otherwise I would be willing and able to pay more"; $25 WTP; 12/06/92 - V1; 4

"$200/yr is all we could afford for what amounts to a non-tax deductible charitable contribution"; $200 WTP; 12/06/92 - V1; 4

"I would be willing to pay that amount and more if I had it. Because the fish and wildlife and my family are important to me."; $200 WTP; 12/06/92 - V1; 6

"This is probably the greatest possible amount I could pay now with a budget already constrained to the max. With my children grown in another few years, possibly I could pay more."; $450 WTP; 12/06/92 -V1; 6

"What I'd be willing to pay and what I can pay are two different things. I'd like to see everything cleaned up, but at the same time, I don't make a whole lot, so I can't pay a whole lot." V.P. 2

"I'm going to take into consideration what I can afford to pay. You've got to remember, do you want this on tape? I'm a widow on a limited income. So that's going to make a difference for what I'm going to put down. When my husband was alive and we were both working..." V.P. E, version 2.

CHAPTER 4
SCOPE TESTING AND EMBEDDING

Drs. Hausman and Desvousges raise a number of issues concerning responsiveness of the WTP values to the scope of the injuries being valued, which is referred to as scope testing. They also raise a related issue concerning embedding in WTP responses and the State's procedures to identify and correct for embedding. The State addressed scope testing issues in Section 5.5 of the CV report and addressed embedding in Sections 2.1, 2.2, and 5.2 of the CV report. Rather than repeat this information, this rebuttal provides additional supporting information about these issues (Sections 4.1 and 4.2) and addresses other related points made by Drs. Hausman and Desvousges (Section 4.3).

Both the original CV report and the more detailed analyses reported in this section demonstrate (a) that respondents were appropriately responsive to scope and (b) that respondents' answers to the disembedding questions were consistent with their responses to other questions in the survey.

4.1 THE SCOPE TEST AND EMBEDDING ISSUES

Scope Test

ARCO's experts focus much of their attack on the supposed failure of the Montana CV study to pass a between-sample statistical test of responsiveness to scope. This test, and others like it, identified by NOAA in proposed regulations [59 Fed. Reg. 1062, January 7, 1994] have now been discarded by NOAA on the basis of extensive negative comments from industry, environmental groups, EPA, and even the members of the Blue Ribbon Panel that NOAA itself commissioned. In fact, the scope test that ARCO's experts view as central to their arguments motivated the contingent valuation panel members' criticism of the NOAA proposed regulations. Four members of the panel (see Arrow, Leamer, Schuman, and Solow, undated memo) wrote:

> The report of the NOAA panel calls for survey results that are "adequately" responsive to the scope of the environmental insult. The proposed scope test is built to assure that there is a statistically detectable sensitivity to scope. This is, in our opinion, an improper interpretation of the word "adequately" . . . A better word than "adequate" ... would have been "plausible": A survey instrument is judged unreliable if it yields estimates which are implausibly unresponsive to the scope of the insult . . . The

> fundamental problem with any purely statistical definition of sensitivity is that it depends (foolishly) on the sample size.

The similarity of scenarios is an important consideration to designing a between-sample scope test. Large, rather than small, differences in scenarios are a more appropriate basis for between-sample scope testing, which is what NOAA had recommended be developed in pretesting. The problem with between-sample testing of similar scenarios is illustrated in the literature (Birnbaum, 1982, discussed below) and in the example already provided in the State's CV report.

> Consider two trees across the street of similar but unequal heights. One sample of individuals is asked to report the height of Tree A in 10 (or 20) foot increments, and a second sample of individuals is asked to report the height of Tree B in 10 (or 20) foot increments. The average of the estimates for both samples may well be the same or very similar, leading to the estimate of the difference in height of the two trees of positive, zero, or even negative value. Yet, if individuals are asked to estimate the difference in heights of the trees when examining both trees together, most individuals will clearly and correctly identify one tree as taller. They can also provide an estimate of how much taller the tree is, and that estimate will generally be much better than the estimate produced using a between-sample difference in results. This example exactly parallels measuring values for environmental cleanup at the Clark Fork sites as conducted in this study and illustrates why the within-sample comparisons in this study are more reliable (pages 5-38).

Dr. Hausman misrepresents the message in this example (pages 21-22). He suggests that even if the relative comparison of (the percent difference in) tree heights is accurate, the absolute measure of the tree heights (in feet) and the difference in tree heights (in feet) are inaccurate and meaningless. This misses the emphasis of the tree example to the between-sample versus within-sample scope tests. Compared side-by-side by the same individual, the individual may measure one tree as about 50 to 60 feet, another tree as about 40 to 50 feet, and the difference as about 8 to 12 feet. Each of these measures has comparable measurement error and are reasonably accurate. Still, in a between-sample comparisons (separate individuals comparing separate trees), the individuals may rate both trees as about 50 feet, or one as 40 feet and one as 60 feet (particularly if given 10 foot response increments to choose from when formulating their response). The absolute measure of height is reasonably accurate, but the between-sample estimate of the difference in heights (0 to 20 feet) is less accurate than when using the within-sample comparison. Again, the within-sample difference is more reliable than the between-sample differences on which Dr. Hausman bases so much of his Montana-specific CV critique.

As we show below, the so-called failure of the Montana CV study to pass the now discredited scope test is the result of scenarios that were designed to conservatively reflect alternative levels of cleanup that were of interest (rather than designed to pass the scope test), and the

use of a sample size chosen not on the basis of conducting a scope test (a notion which had not even been proposed at the time of the study), but rather to be sufficient to identify the magnitude of damages (see the end of Section 4.2 for additional discussion of sample size issues). In spite of these considerations, the study passes such a between-sample scope test if one uses appropriate samples, as discussed in Section 4.2.

Embedding

The issue of scope arose in the considerations of the NOAA CV panel because of a concern over the issue of embedding. In contrast to the vast majority of the refereed scientific literature on CV, which consists of literally hundreds of papers, the issue of embedding has been addressed only in a handful of papers. The approach taken by the vast majority of CV practitioners has been to ignore embedding or to naively assume one could make it go away by changing the survey design. The authors of the Montana CV study have been among the few practitioners to apply a relatively new and conservative approach to deal with this potential problem. Since the embedding problem supposedly arises because some respondents include more in their stated value than the researcher intends (e.g., values for a more extensive and inclusive list of environmental commodities), the Montana CV survey challenges respondents' values by reminding them of other environmental commodities and good causes upon which they might wish to spend the dollar value they assigned to cleaning up the Clark Fork sites. The question then asks them to revise their WTP response downward if appropriate.

Drs. Hausman (pages 34-36) and Desvousges (pages 17-20) conclude that the State has failed in its CV study design because it did not design embedding out of the survey, and they complain that the State has to adjust the initial WTP to account for this issues. This is a naive presentation. Embedding can exist in WTP responses even when the survey designer hopes and believes he has eliminated it. The State's position is that it is better to presume one has not entirely omitted embedding, and that it is better to over-adjust for possible embedding by reducing values when the reduction may be simply associated with an overlapping motive for cleaning up the upper Clark Fork sites tied to a general desire to enhance the environment (discussed below). Also, the State does not adjust the WTP values, the State allows the respondents to revise their responses, and it is the respondents' revised responses that are reported. We argue that the embedding-adjusted WTP values used in the damage assessment are conservative and reliable for at least the following reasons:

- First, the comparison studies discussed in Section 2.1 of this report, which have validated the use of traditional CV, did not employ such a procedure. Since the procedure lowers values by approximately 50 percent, it is highly conservative.

- Second, the NOAA Panel on CV recommended that respondents be reminded of substitutes (other commodities which might serve similar purposes). The

disembedding question reminds people of other uses for their money in one of the most forceful and direct ways possible.

- The third reason has to do with people's motives for cleaning up the upper Clark Fork sites. When asked if their WTP is just for the upper Clark Fork sites, many respondents can truthfully indicate that the WTP is partially motivated by a specific desire to clean up the upper Clark Fork sites, and partially motivated by a general desire to enhance the environment, which cleaning up the upper Clark Fork accomplishes (e.g., responses 2 and 3 to the embedding Question 30). Then, for example, a response to Question 31 that 50 percent of their WTP is just because its the upper Clark Fork, and 50 percent because the upper Clark Fork cleanup is a good environmental cause is entirely consistent with communicating their motives for paying to cleanup these sites. When this occurs, correcting for embedding by reducing the WTP by 50 percent is conservative because some or all of the other 50 percent is for the upper Clark Fork cleanup as a reasonable alternative to achieve general environmental goals.

- Fourth, experimental economists have shown that people tend to lower their initial values as they gain more experience with valuing a commodity (see the text on experimental economics by Davis and Holt, 1993). Thus, it should be completely unsurprising that respondents will generally revise their bids downward given the opportunity to reflect.

- Respondents reflect on and make their own adjustments to their responses. The State's CV instrument and analysis facilitates, records, and analyzes the respondent's answer.

Each of these reasons suggest that the approach used in the Montana study to deal with potential embedding problems is likely to produce conservative estimates, especially compared to traditional CV studies.

4.2 ANALYSIS

We now turn to additional analysis of embedding and scope within the Montana CV study. Dr. Hausman contends that the State is incorrect in its claim that the within-sample scope test is preferred to the between-sample scope test. He bases his claim on a statistical test showing that the estimates obtained from the two types of comparisons are significantly different. Even if we grant for the moment the validity of the statistical test employed by Dr. Hausman, there is no *a priori* reason to believe that it is the within-sample, and not the between-sample, estimate that is biased. In fact, in the published scientific literature there are ample demonstrations that in many judgment situations it is the between-sample, and not the within-

sample estimate, that is biased or nonsensical. For example, Birnbaum (1982) reviews a number of such instances and concludes:[1]

> It is not unusual in a between-subject design for the smaller stimulus to be judged greater. For example, . . . the 27 dot pattern (in the positive context) is judged darker than is the 40 dot pattern (in the negative context). However, *within* each context (i.e., within subjects), the 40 dot stimulus is always judged darker. Therefore, comparing judgements of stimuli between groups of subjects who experience different context can be misleading. (page 445).

Nickerson and McClelland (1989) illustrate that between-sample tests are expected to be most misleading when they aggregate across people with different judgment policies. As we show below with respect to the embedding issues, there are four distinct types of respondents in the State's survey. Understanding these different types of respondents provides sound reason for preferring within-sample to between-sample estimates in this context.

There are four types of respondents in this survey, who identified themselves by their responses to Q30. The four types of respondents can be characterized by their types of concerns, which are (1) *site-specific*, (2) *site-specific* and *generalized*, (3) *generalized*, and (4) *idiosyncratic*. These types correspond to the four response alternatives to Q30.
Dr. Hausman contends that by so answering Q30, respondents are admitting to a mistake in responding to the WTP question (Q28). However, this is definitely not the case because the respondents in these four concern groups responded to Questions 28 through 31 consistently with their other responses throughout the survey.

Consider first group 1 with *site-specific* concerns. On the very first question these respondents were significantly more likely to be aware of the upper Clark Fork NPL sites than those in groups 3 and 4 (and more than those in group 2 in Version 2 of the survey but not Version 1). Furthermore, as would be expected, they were *not* significantly more likely to be aware of the other four NPL sites not involved in this case. On Q2, those in the group with *site-specific* concerns generally rated the importance of other issues as less important than those in other groups. Even after reading more about the four other NPL sites, the *site-specific* group persisted in rating other sites as less important than do those in the other groups. In summary, even before identifying themselves in the disembedding question (Q30) as having *site-specific* concerns, respondents in this group, relative to the other groups, were more likely to be aware of the sites, less likely to care about other general problems, and less likely to care about the other four NPL sites. When these respondents report that all of their stated value in Q28 is to

[1] The context of the quoted example was a psychophysical experiment in which people judged the darkness of squares filled with varying numbers of dots, the more dots the darker the appearance of the square. Side-by-side, the 40-dot stimulus is judged darker than the 27-dot stimulus, but when viewed independently in different distributions (either positively-skewed or negatively-skewed around the same mean) of stimuli, the 27-dot stimulus is judged darker than the 40-dot stimulus, on average.

be applied to the four upper Clark Fork NPL sites, there is every reason to believe them because their responses have been consistent with this position throughout the survey. There is *no* reason to believe, as Dr. Hausman speculates without data or other support in footnote 21 on page 35, that these respondents with *site-specific* concerns either "unconsciously embedded" or were "unwilling to admit" to a "mistake".

Now consider group two with both *site-specific* and *general* concerns. They were, relative to the other groups, likely to be aware of the upper Clark Fork sites but not more likely to be aware of the other four NPL sites. This establishes their site-specific concern at the very beginning of the survey. However, they were also concerned with other problems such as education, air pollution, jobs, roads, etc. (in Q2). After reading about the other four NPL sites, they tended to rate these sites more importantly than did those in group 1. This establishes, especially relative to the *site-specific* group, that this second group also has general concerns. Finally, those in the *site-specific* and *general* concerns group felt in general more responsible (Q34) than those in the other groups for the Clark Fork sites. Thus, they were not making a "mistake" when allocating, on average, 50 percent of their WTP response to the Clark Fork sites and 50 percent "to partly clean up other hazardous waste sites." Instead, they were being entirely consistent with the views and preferences they had already stated in the survey.

Next, consider group three with primarily *general* concerns. This group, relative to the others, tended not to be aware of the upper Clark Fork sites, to care the most about general problems, and not to feel particularly responsible for the upper Clark Fork sites. Thus, while the upper Clark Fork sites are part of their general concerns, they do not particularly care about those sites and hence allocate a smaller portion of their WTP to those sites. In other words, their response to the disembedding question is entirely consistent with their prior responses in the survey.

The fourth group and smallest group has *idiosyncratic* concerns. This group, relative to the others, was least aware of the upper Clark Fork NPL sites, was least concerned about other problems, and after reading about the sites was still the least concerned and felt the least responsibility. Not surprisingly, then, 50 percent of this group had adjusted WTP values of $0.

Once it is recognized that there were four conceptually distinct groups of respondents to the survey, it is easy to see how averaging results across all four groups could produce very misleading between-sample comparisons as indicated by Nickerson and McClelland (1989). Clearly, a between-sample scope difference should only be expected for those respondents with a high degree of site-specific concerns, namely groups 1 and 2. As shown in Table 4-1, those two groups do show a between-sample scope difference. The difference is statistically significant for group 2 (t statistic = 1.77, one-tailed p = .04), even with the small sample sizes, and is approximately equal to the estimate used to calculate the value of the residual

Table 4-1
Between-Sample and Within-Sample Comparisons of the Difference in WTP for Complete and Partial Cleanup of the Upper Clark Fork NPL Sites

Sample Group: Q30 response and descriptor	Version	Sample Size	Between-Sample Comparisons: Mean WTP[1] (SE of Mean)	Within-Sample WTP Difference by Version: Mean Difference[2] (SE of Mean)
1 = Site Specific Concerns only	1	56	$70.77 ($7.82)	$21.63 ($5.00)
	2	58	$61.47 ($7.81)	$20.43 ($4.78)
	Difference		$9.30 ($11.06)	
2 = Site Specific and General Concerns	1	119	$55.20 ($5.23)	$16.58 ($2.34)
	2	120	$42.35 ($5.01)	$15.33 ($5.86)
	Difference		$12.85 ($7.24)	
3 = General Concerns	1	140	$29.24 ($3.62)	$13.56 ($2.39)
	2	151	$32.13 ($4.21)	$13.50 ($5.44)
	Difference		-$2.89 ($5.59)	
4 = Idiosyncratic Concerns	1	29	$11.93 ($3.77)	$4.91 ($2.33)
	2	23	$26.74 ($10.18)	$2.01 ($2.44)
			-$14.81 ($9.99)[3]	

[1] The within-sample comparisons list values for the first WTP asked:
- Version 1: Adjusted WTP for complete cleanup
- Version 2: Adjusted WTP for partial cleanup
- Difference: Version 1 complete cleanup WTP minus Version 2 partial cleanup adjusted WTP.

[2] The between-sample differences list, for respondents to each survey version, the difference in adjusted WTP for complete cleanup minus adjusted WTP for partial cleanup.

[3] The very different standard errors make the pooled standard error questionable and invalidates the usual t-test.

damages in the original report.[2] The between-sample scope difference is approximately the same dollar amount as in the within-sample comparisons, but is not statistically significant using conventional significance levels. In contrast, we would *not* expect those with *general* concerns to be sensitive to scope in a between-sample comparison. Indeed, as expected, the difference for group 3 is small, in the wrong direction, and is not close to statistical significance. For group 4, with *idiosyncratic* concerns, adjusted WTP was $0 for 50 percent of the respondents, so the means are determined by just a few individuals and are not statistically stable. (For example, a few more large WTP values happened to be in Version 2.)

Note that, due to the idiosyncratic concerns of group 4, there would be good scientific grounds for omitting them from the sample. NOAA recognized this issue when it first proposed the between-sample scope test.

> The issue is complicated by the possibility . . . that a significant minority of the population may be insensitive to any reasonable differences in scenarios: some individuals may not be willing to pay anything for any environmental cleanup, others may be willing to pay unrealistically high (and invariant) amounts for any size environmental cleanup. The trustees are to develop procedures for identifying these people, so that the demonstration that the scenarios are meaningfully different would rest on the remaining participants . . . . NOAA is seeking comments on . . . how should "insensitive" individuals be excluded? (page 1146 column 1)

To apply the between-sample comparison test to group 4 respondents amounts to measurement and statistical trickery. The test requires these respondents to separately (rather than side-by-side) assign values that are statistically significantly different for two levels of cleanup they valued as similar for sites where they have basically general concerns. Omitting such individuals in the State's CV study would have the effect of increasing the estimate of residual damages; thus, as a conservative procedure, group 4 respondents were included in calculations of damage estimates.

The data in Table 4-1 reflect many other notable consistent response patterns in the CV data. First, as site specific concern decreases (from group 1 to group 4), the WTP also decreases. This is true for WTP for complete and partial cleanup (the fourth column), and the difference in WTP (the fifth column). These results are completely consistent with the respondents' answers throughout the survey. Second, note that, for each group, the two alternative within-sample estimates of the difference between partial and complete WTP are virtually identical. In other words, for each group, the same within-sample difference is obtained regardless of whether respondents start with valuing complete cleanup or partial cleanup.

---

[2] The between-sample difference for Group 1 is of the correct sign and magnitude but is not statistically significant because of the small sample size relative to the variance in WTP across individuals.

Dr. Hausman claims that the sample size of the State's CV study does not help to explain the failure of the between-sample scope test (page 28). However, when one understands and accepts the difference in respondent types, Dr. Hausman's comment is found to be invalid. With substantial increases in sample sizes, the between-sample difference in WTP for groups 3 and 4 is expected to move to zero (rather than the current findings that WTP is larger for partial cleanup than for complete cleanup), and the between-sample difference in WTP for groups 1 and 2 will become increasingly statistically significant such that the sample wide between-sample difference in WTP will become statistically significant.

In conclusion, it is important to reiterate that the State's cleanup scenarios and sample sizes were designed prior to the issuance of any proposed between-sample scope tests. Consequently, they were not designed to pass these artificial scope tests (this issue is discussed at length in Section 5.5 of the CV report). Nevertheless, the results indicate that for those respondents who have *site-specific*, and *site-specific and general* concerns (groups 1 and 2), the between-sample difference in WTP values have the expected sign and magnitude and these differences are consistent with the size of the more accurate within-sample differences in WTP. Those individuals who only have *general* concerns (group 3), and for whom complete cleanup and partial cleanup at these specific sites may not be seen as much different, do not show between-sample responsiveness to changes in the scenario, although they report within-sample responsiveness. This reflects that they have more difficulty with the absolute measurement of values than the relative measurement of values for issues of lesser concern. As the psychological literature indicates, within-sample comparisons are more appropriate and more accurate for these individuals.

4.3 OTHER RELATED ISSUES

Drs. Hausman and Desvousges make a number of other miscellaneous comments about embedding, scope testing, and the survey results that are misleading and incorrect. Several of these other comments are briefly addressed in this section.

The State's CV Study was not Designed with Between-Sample Scope Testing in Mind

Dr. Desvousges contends the State designed its study with two versions and says this was done "implying a plan to test of some type of difference in versions" (page 37 footnote 31) and "with the intent of comparing those results" (page 37 footnote 32). As stated in the State's CV report, and in depositions, two versions were undertaken to examine survey design issues related to the order of asking the questions on the difference in the values for complete and partial cleanup (State CV report page 2-1, and the emphasis throughout the report on the computation of residual value). The State did not design its CV study for the NOAA between-sample scope test, which was issued after the survey was designed and implemented.

Exactly Equal Satisfaction Ratings are not Required to have Similar Values

In Section 5.5.3 of the State's CV report, the State notes that 44 percent of respondents assign very similar ratings to their satisfaction with the partial cleanup and complete cleanup scenarios, even when these two scenarios are compared side-by-side (NOAA's proposed test suggested that different scenarios be designed so that at least 5 percent rate the scenarios as the same). This illustrates that the scenarios are too similar to pass a between-sample scope test. For this analysis, the State considered rankings of partial cleanup and complete cleanup that, on a one to seven scale, are within one unit of each other as being very similar in that the differences in values would be small.[3] Dr. Desvousges takes this analysis to the extreme by suggesting that people can only have, or report, similar values for the two cleanup scenarios if they rate their satisfaction with the two scenarios as exactly equal (pages 39-40). This clearly is an unrealistic standard. On the other hand, Dr. Hausman suggests that large differences in respondent rankings for partial and complete cleanup are inconsistent with the WTP difference of $14 (pages 29-30). Quite the opposite is true. The sum of the partial cleanup ranking scores are 42 percent less than the sum of the complete cleanup ranking scores. The within-sample difference in WTP, in Version 1, is 35.6 percent. These findings are consistent. Dr. Hausman also seems to imply that the difference in rankings should be associated with larger dollar differences, although this contradicts his earlier statement that he feels the WTP values are implausibly large (pages 24-26). There is no *a priori* basis to assert that a point value on the ranking scale should be associated with a specific dollar amount.

Finally, Dr. Desvousges notes that the sample wide distributions of satisfaction ratings for complete cleanup and partial cleanup are statistically different (pages 39-40). This is not an indictment of the State's CV survey or analysis. It simply confirms our results in Table 4-1; while there are a substantial number of individuals who rate, and value, the two cleanups as very similar, there also are a substantial number of individuals who rate, and value, the two cleanups as very different.

Anchoring and Sequencing Bias

Dr. Hausman suggests the Montana survey results support his general claim for anchoring and sequencing bias in CV surveys in general (pages 11). On page 32, Dr. Hausman specifically

---

[3] Summed across the three resource groups, the average difference in ratings between partial and complete cleanup is 6.93 points (out of 21 possible points). The mean ratings of partial and complete cleanup are: 3.10 and 5.68 for aquatic resources, 3.19 and 5.40 for groundwater resources, and 3.38 and 5.53 for upland resources (State's CV report, Table 4-3; individual questions use a seven point scale). The average within-sample difference in WTP for complete and partial cleanup is $14. If one assumes the rating scale and valuation can be exactly lined-up in a linear fashion (a one unit change in rating is associated with a constant dollar value, which is an assumption we do not make), dividing $14 by a rating difference of 6.93 amounts to $2.02 per rating point - or a small value difference in terms of between-sample scope testing.

indicates that "In response to the attitudinal questions concerning satisfaction with complete and partial cleanup, a number of respondents indicate that they would be more satisfied with partial cleanup than with complete cleanup. . . However, for a number of respondents who stated that they preferred partial cleanup to complete cleanup, the WTP answer for complete cleanup exceeds the WTP answer for partial cleanup. This result demonstrates that the questionnaire induced an anchoring or sequence effect" (page 32).[4]

Dr. Hausman's conclusion that these results imply anchoring and sequencing effect of any significance is unfounded. Only three percent of the sample exhibited the preference reversal response pattern identified by Dr. Hausman, while 97 percent of responses do not show this preference reversal.[5] Even thought this analysis requires the exact comparison of an attitudinal scale results with monetary scale results, the overwhelming majority of respondents show consistent responses. In a general population survey, a few such anomalous results are normal when comparing different types of scales.

Combined with the discussion in Section 4.2 above, and the results in Table 4-1, which indicate that within-sample comparisons are consistent and are more reliable for all sample groups, and consistent with between-sample differences for individuals in Groups 1 and 2 (Table 4-1) who hold site-specific concerns and can be expected to have more precise WTP responses, the data show that Dr. Hausman's conclusions of anchoring and sequence biases are unfounded.

Interpreting the Embedding Question Responses

Dr. Hausman makes another particularly misleading presentation when discussing the embedding question. He notes (page 36) that respondents who responded to Question 30 that their response was "Basically a contribution for all environmental and other causes" gave an unadjusted (for embedding) WTP response of $74.84. Adjusting for embedding, the WTP for just the upper Clark Fork Basin sites is about $29.24. From this he states "In other words, to cleanup the Clark Fork sites, these respondents are willing to pay $29, but to additionally address all other causes (presumably all environmental, social, political, etc.), they are willing to pay only an additional $46. This result is simply economically implausible . . ." What is implausible, and incorrect, is Dr. Hausman's interpretation of these results. Recall that respondents were asked for their WTP for cleaning up the upper Clark Fork sites (Question

[4] Dr. Hausman also cites Tversky and Kahneman (1974) and Green et al. (1995) to support his general claim of anchoring effects (pages 11 and 36). The first paper deals with an entirely different type of anchoring effect than Dr. Hausman is discussing. The second paper is an unpublished university "mimeo" that was not provided by Dr. Hausman in discovery.

[5] We define a preference reversal as when the sum of the satisfaction scores for partial cleanup (Questions 4, 8, and 12) is larger than the sum of satisfaction scores for complete cleanup (Questions 4, 8, and 13), but WTP for complete cleanup is larger than WTP for partial cleanup.

28), asked to comment on their answer (Question 29), told that "it is difficult to think about paying to clean up just one site or even just one environmental problem", asked in Question 30 to indicate whether any or all of their WTP may be motivated by other concerns (which may or may not overlap with the concern about cleaning up the upper Clark Fork sites), and finally asked to state what percent of their WTP is just for cleanup of the upper Clark Fork sites (Question 31). The Hausman interpretation of these results, that the $79 WTP is for all causes and results in the $29 for cleanup of the upper Clark Fork sites being disproportionate to the $46 for all other causes, is inconsistent with the questions that were asked and is inconsistent with a common sense interpretation of how respondents answer these types of questions.

The more appropriate interpretation follows the discussion at the end of Section 4.1 above. That is, these respondents are indicating that their WTP is motivated in part by a specific support of the cleanup of the upper Clark Fork sites (the $29 value), and motivated in part by a general desire to clean up the environment, which cleaning up the upper Clark Fork sites also accomplishes (the $46 value, which they would be equally willing to apply to this cleanup or to another problem). This response to the embedding question is entirely consistent with the other survey responses by these group 3 individuals that their stated value *for cleaning up the upper Clark Fork sites* is motivated more by a general desire to enhance the environment in Montana than by a specific concern for this specific site.

Dr. Hausman also questions whether the embedding question will sufficiently adjust for embedding in the WTP responses (page 35). The above discussion and analysis indicates that the embedding question may instead be over-adjusting for embedding by eliminating that part of the WTP for cleanup of the upper Clark Fork sites that is motivated by general environmental concerns as much as it is motivated by site-specific concerns.

Dr. Desvousges notes that 78 percent of respondents indicate that their responses are not just for the cleanup of the upper Clark Fork sites, and this is a significant flaw of the study (page 20). The number of people who adjust their WTP response in Questions 30 and 31 is not a measure of reliability. This percent is influenced by the separability of the good from other goods and services, the degree to which this cleanup will meet general environmental cleanup goals (motives), and the population considered (we considered all Montana residents rather than just local residents). It also reflects the influence of a reminder about substitutes and income constraints (Section 4.1). The percent of respondents who adjust their response following these questions indicates the significance of the follow-up question, as done in this survey, rather than ignoring embedding and substitutes (including the assumption that one has fully eliminated embedding in the WTP question design) as in most CV studies.

Dr. Desvousges and Dr. Hausman repeatedly suggest that respondents are responding with a "warm glow" value that does not reflect damages (Dr. Desvousges at page 20 and Dr. Hausman at page 14). The warm glow phenomenon is intended, apparently, to refer to WTP provided for the pure enjoyment of giving money to causes. As a critique of CV responses,

this catchy phrase has little merit. No evidence has been provided that people respond to WTP questions just for the sake of handing over money. Second, paying to clean up the upper Clark Fork sites because they represent the types of general environmental efforts a respondent supports is no more invalid than purchasing a specific piece of art while also stating one likes to purchase art. The price paid reflects both a general desire to purchase art and the specific WTP for the specific piece of art.

Dr. Hausman's Claim of Data Mining

When discussing the State's investigation into the scope test by examining the sample partitioned by their response to Q30 (the embedding question), Dr. Hausman states the State was "data mining" and "data snooping," which he claims invalidates the analysis (page 33 where he refers to the Rowe deposition, at page 69). This statement is particularly egregious as it was made clear in the Rowe deposition that this analysis was designed and conducted, as presented, because this same investigation was conducted in two previous CV studies where the same embedding question was asked (Rowe deposition at pages 71-72).

CHAPTER 5
SCENARIO DEFINITIONS

ARCO's experts contend that the commodity valued in the Montana survey on the basis of the scenarios presented respondents was improperly defined and characterized. Again, these arguments are simply incorrect or misleading.

5.1 SUBSTITUTES

ARCO's experts argue that the Montana CV survey did not adequately remind respondents of the availability of substitutes (Hausman pages 33-38 and Desvousges pages 23-25). Their arguments rely on an incorrect interpretation of the term substitutes which is inconsistent with elementary economic theory. In his basic economics text Dolan (1977) defines substitute commodities as (page 41):

> "A pair of goods for which an increase in the price of one good causes an increase in the quantity demanded of the other, other things being equal."

For example, if the price of Pepsi increases, the demand for Coke will increase because the two goods are substitutes. Thus, if an increase in the price or cost of any good increases the demand for natural resources (and consequently, cleanup) of the upper Clark Fork sites then that good is a substitute for the resources of the upper Clark Fork. Clearly, cleanup of other sites that restore natural resources may be a substitute for cleanup of the upper Clark Fork sites if the natural resources are themselves substitutes. Both Drs. Hausman and Desvousges contend that cleanup of other sites is not a substitute for cleanup of the upper Clark Fork sites. Although this is a purely empirical question, it is logical to assume that similar natural resources (e.g., fish, wildlife) at different sites are likely to be viewed as substitutes for natural resources at the upper Clark Fork sites. Further, it would be incorrect not to inform respondents of the need to clean up other sites, which might compete for their cleanup dollars, thus lowering values for cleanup of the upper Clark Fork sites. In fact, when respondents are reminded of other sites in the embedding question (Q30), values are lowered by about 50 percent. Thus, the data indicates that Drs. Hausman and Desvousges are incorrect in their argument that cleanup of other sites does not constitute a substitute good.

Further, Dr. Desvousges argues that the remainder of the availability of uninjured substitute resources (contained in Questions 6, 10, and 14 of the survey and incorporated specifically to remind respondents of the availability of uninjured resources) that could instead be purchased by the State "cannot be considered a substitute resource" because "the resource is neither available nor accessible to the public until the State acquires it" (page 25). Clearly, this is

incorrect. These alternative resources are substitutes for the injured resources (just as Coke is a substitute for Pepsi) because there are opportunities to acquire the resources (or Coke) regardless of whether the resources (or Coke) is actually acquired. Again ARCO's experts' arguments are inconsistent with fundamental concepts of economics.

Finally ARCO's experts assume a remarkable degree of ignorance on the part of Montana's citizens who, in fact, are well aware of the rich availability of unspoiled natural resources in the State.

As indicated in the State's CV report (page 1-8), reminders of substitute commodities and alternative expenditure possibilities in the survey include maps of Montana, which indicate the confined geographic area of injuries; listings of alternative public policy issues facing Montana residents; questions on alternative NPL sites in Montana; questions on alternative cleanup options at the upper Clark Fork NPL sites; and the final reminder in the embedding question (Question 30). Substitutes for the cleanup of the upper Clark Fork sites are addressed.

5.2 COMMODITY DEFINITION

Dr. Desvousges contends that the commodity used in the survey was incorrectly defined (page 26). He first argues that a subcommittee of the U.S. EPA Science Advisory Board, the Environmental Economics Advisory Committee (EEAC) was critical of an unrelated study. He fails to mention the U.S. EPA, in effect, rejected the EEAC report (see memorandum from Barnes Johnson, Chief Regulatory Analysis Branch, U.S. EPA, October 20, 1992, and letter of Carol Browner, Administrator, U.S. EPA, March 21, 1994). Dr. Desvousges further argues that the CV survey misrepresents baseline conditions (page 27), but offers no support for these opinions. Both Dr. Hausman (page 37) and Dr. Desvousges (page 29) argue that respondents should have been informed of quantitative rather than relative measures of environmental quality (e.g., fish per mile, deer per square mile). In fact, respondents often find such numerical descriptors difficult to understand. Rather, they have a perception of normal levels of fish in unpolluted streams, which they have experienced, and generally find relative measures, such as a 75 percent reduction in stocks, to be more meaningful.

Finally, in representing baseline conditions, care was taken to show respondents photos that appropriately represented baseline streams. Similarly, the upper Clark Fork sites' proximity to urban areas, highways, etc., was depicted on the map enclosed with the survey instrument.

CHAPTER 6
WTP ELICITATION

Drs. Hausman and Desvousges critique the State's WTP elicitation procedure by focusing on the use of a 10-year payment period, the payment question, the types of payment vehicles used, and the use of a payment card. They infer that the State's design results in unreliable and upwardly biased WTP responses. The literature and the State's CV survey data do not support their conclusions.

6.1 TEN-YEAR PAYMENT PERIOD

Drs. Hausman (pages 39-40) and Desvousges (pages 45-47) suggest that WTP should have been measured using a one-time, or lump sum, payment rather than an annual payment over 10 years. A lump sum payment would result in a more conservative, but not more accurate, damage estimate that may be biased downward.

In its response to the NOAA and DOI proposed CV rules issued in early 1994, the U.S. EPA (1994) recommended that there should be no preference for lump-sum payments, and we concur. The EPA comments indicate that forcing respondents to estimate what they would pay as a one-time payment right now to obtain that benefit over an extended period of time could be expected to result in an understatement of the stream of benefits. This occurs for several reasons, including the following:

- Uncertainty about the future could cause the individual to discount heavily potential benefits in future years. If individuals have uncertainty about whether they will be in a location to receive future benefits of a program, or about whether a program will succeed, they will naturally reduce the up-front lump-sum payment they are willing to pay to receive future benefits. This biases the measure of value of the actual cleanup benefits downward.

- Current income and credit constraints may limit the perceived ability to pay an amount equivalent to the present value of an annual stream of payments. A lump-sum payment may create short-run budget constraints for some respondents and restrict what they are able and willing to pay in one period compared to what they are able and willing to pay through time. In fact, some

Table 6-1
Comparisons of Values Measured Using WTP Questions with Dichotomous Choice Questions

Study	Commodity	Value from Dichotomous Choice Question/Value from WTP Question
Seller, Stoll, Chavas (1985)	Recreational Boating	
	WTP for Lake Conroe	5.9
	WTP for Lake Livingston	4.8
	WTP for Lake Houston	9.5
	Average	6.7
Boyle, Bishop (1988)	Scenic Beauty	3.1
Johnson, Bregenzer, Shelby (1990)	Noncommercial River Rafting	1.6
Walsh, Johnson, McKean (1992)	Outdoor Recreation	
	WTP for activity day	1.3
Kealy, Turner (1993)	Acid Deposition/Chocolate Bar	1.4 - 2.5
	WTP for Public Good Average	1.95
	WTP for Private Good	1.0
McFadden (1994)	Preservation of Rocky Mountain Wilderness Areas	
	WTP for 1 area	11.9 - 72.9
	WTP for 57 areas	10.2 - 41.1
	Average	34
Summary	Range	1.0 - 72.9
	Average	7.0
	Median	1.9

vehicles). ARCO's experts raise several concerns with the payment vehicles used in the State's study. These concerns are speculative, unsupported by the data, and misleading.

Dr. Hausman first suggests that respondents might respond strategically to influence the policy makers' comparison of benefits and costs to arrive at the respondent's desired outcome (pages 41-42), a phenomenon referred to in the literature as "strategic bias." He asserts that the net impact of strategic bias is to inflate the computed damages. Several studies have tested for such strategic behavior and it has not been found to be a significant issue (Rowe et al., 1980; Mitchell and Carson, 1989; EPA, 1994). This concept seems to be one that best reflects the thoughts of economists and litigants and not the thoughts of citizens responding to CV surveys.

Next, Dr. Hausman suggests that respondents will select payment vehicles where they would not have to bear any of the costs of cleanup and, because they would not pay, "they have a further incentive to exaggerate their WTP answer" (page 42; Desvousges, pages 47-49, provides a similar comment). This is in stark contrast to the State's CV survey results (CV report Chapter 5).

Just over 80 percent of respondents choose one or more payment vehicles that require that they participate in paying for the cleanup (e.g., increasing prices and taxes you pay). Only 20 percent chose only "other payment vehicles" where they may or may not have to pay (e.g., depending on whether the respondent hunts, fishes, or drives down I-90). The WTP by those who chose only "other payment vehicles" is approximately 14 percent (Version 1) to 26 percent (Version 2) of the WTP of the remainder of the sample. These individuals more frequently write protest comments on the CV survey and have a lower self-reported responsibility to pay for site cleanup compared to the rest of the respondents. This suggests that the WTP responses by individuals who select only "other payment vehicles" are more likely to have scenario rejection that leads to downwardly biased WTP. Even if one accepted an extreme interpretation of the Hausman hypothesis — that all of the respondents who cited only "other payment vehicles" had a true WTP of $0 and their reported WTP are amounts they thought others should pay (and not themselves) — the net impact on the sample-wide WTP would be less than 5 percent. Clearly, the validity and significance of this issue are misrepresented in the ARCO comments.

Dr. Desvousges also raises concerns about the payment vehicle of "industry pays with some costs being passed on to consumers in higher prices" (page 47). He asserts that "it is likely that respondents viewed this payment vehicle as allowing them to shift the burden to industry." Our experience is that a large share of respondents see this as sharing (not shifting) the responsibility, which reduces vehicle protest. This is consistent with the wording of the relevant questions in the CV survey, which stress the amount your household will pay, not

others.[2] Respondent comments indicate that those who see this payment vehicle as strictly shifting the burden to industry are likely to provide a $0 WTP or a reduced WTP as a form of protest response. As identified in the State's CV report (pages 2-10, 2-11), payment vehicle problems lead respondents to underreport their values for cleanup because they don't like the method of payment. Therefore, the Desvousges comment that some vehicles "allowed them to avoid paying anything" is actually associated with behavior that leads to a downward bias in estimated values. The Montana survey approach reduces this problem by seeking vehicles that are acceptable to the respondent. This is a strength of the study, not a weakness.

One confusion raised by Desvousges (pages 20-21) concerns respondents who make comments to the effect of "if everyone paid this, we could pay for the cleanup," a phenomenon he labels "cost calculating." Dr. Desvousges claims this response "severed the link between willingness-to-pay and services" (page 20) and "these values do not represent economic values" (page 21). However, simple logic shows that such WTP responses will be equal to or *less than* the individual's value for the services they will receive from the cleanup. The stated WTP does not exceed the individual's value because if it did they would not be willing to pay this amount. The stated WTP may be less than the individual's value when the individual feels there is not a need to pay their maximum value because if everyone paid this amount, the revenues would cover expected costs. In summary, there is nothing wrong with respondents considering costs, income, and prices of other goods and services when determining their WTP. This is consistent with economic decisions for market goods, political referenda, and nonmarket goods. In the Montana CV study, when "cost calculating" occurs it does not sever the link between WTP and value, but it may cause downwardly biased measures of value.

Dr. Desvousges (page 45) also makes a comment that one cannot adjust the Montana CV results for the payment vehicle selected. Our response is that the only need to make such an adjustment is to remove downward payment vehicle bias, if any. This is not done in the State's analysis and results in a downward bias in the estimate of damages.

In conclusion, the evidence in the literature and in the State's CV survey is that very few, if any, respondents respond, as suggested by the ARCO experts, with convoluted computations to influence results or to report what industry or other individuals should pay on their behalf.

---

[2] The payment vehicle selection questions (Q27) lists the first payment vehicle option as: "Increase in waste disposal fees and taxes paid by industry. Industry would pass along some of these costs to you in higher prices for all goods and services you buy." The second, third, and fourth options all end with the words ". . . bills (taxes) you pay," again emphasizing payments by the affected household.

The WTP question (Q28) asks "What is **the most** your household would be willing to pay . . . through the methods you selected in Question 27 . . . ." No question asks what the respondent thinks others should pay. As provided in the State's CV report (Table 5-26), the written comments by respondents show ample evidence that respondents are concerned with the amount they are paying.

6.2.3 Payment Card Issues

ARCO's experts assert that the payment card used in the State's CV study is flawed as a result of measurement error and because of potential range and centering biases. These comments are misleading and inconsistent with peer-reviewed empirical results based on a test for these biases with the payment card design used in the State's study (Rowe, Schulze, and Breffle, forthcoming). The ARCO experts also comment that the payment card does not include a "No Answer" category and that the $20 incentive fee for completing the survey may amount to a bias. Below we show that these comments are inconsequential.

In the State's CV survey, respondents are asked to respond to the WTP question (Q28) by circling the maximum amount "you would pay annually for 10 years" from a payment card listing 23 different responses ranging from $0 to "more than $5,000.

Measurement Error

Dr. Desvousges (page 49) begins his comments on the payment card design by raising a seriously misleading concern with the "absolute" level of potential measurement error in the payment card. Dr. Desvousges gives an example of an individual with a "true" value of $2,800. The nearest payment card options are $2,500 or $3,300. Selecting one of these two options could result in a $300 or $500 measurement error depending on which option the individual chooses. This is an interesting example as Dr. Desvousges knew from the State's report that after data cleaning, only 3 of 774 first bid WTP responses are in excess of $450 (1 at $650 and 2 at $1,000) and that prior to the embedding adjustment over 70 percent of respondents report first bid WTP values under $100 ($40 is the median response). After correcting for embedding, in Version 1 for example, over 70 percent of responses are less than $50 (the median is $18 and only 11 percent exceeded $100). Therefore, the absolute magnitude of errors suggested by the Desvousges' example is over 100 times larger than the median measurement error in the values used to compute damages.

Dr. Hausman (pages 43-44) makes a related point, stating that a plot of the WTP distribution does not match a lognormal distribution, suggesting there is substantive measurement error. The first problem with this comment is that neither the State nor Dr. Hausman provides any opinion or basis for why, if at all, the WTP distribution should be lognormal (and in the extreme why it should perfectly fit a lognormal distribution).[3] Second, he notes that the number of observations at values such as $10, $60, and $125 is more than he would predict with his lognormal function. As addressed in Rowe, Schulze, and Breffle (forthcoming), this results from respondents rounding to key values and from the breaks in the payment card. This does not create either an *a priori* upward or downward bias.

[3] The state does suggest that measurement errors in the WTP responses may be distributed approximately lognormally. This does not imply that the measured WTP values are necessarily lognormally distributed.

Range and Centering Biases

Dr. Desvousges (pages 49-51) claims that payment cards, used to record respondent values, induce "range and centering biases." His claim is based on references to authors who speculate on this issue but who do not provide solid empirical evidence. A peer-reviewed empirical test for range and centering biases in payment cards provides solid empirical evidence that refutes ARCO's claims (Rowe, Schulze, and Breffle, forthcoming). ARCO was provided a then-current draft of this paper in discovery and chose neither to acknowledge nor to refute these findings in formulating its critique. Testing the same payment card design used in the State's CV study, for the application of WTP for cleaning up hazardous waste sites, the paper concludes that neither range bias nor centering bias is supported.[4] The paper tested four versions of a payment card with the same design as in the Montana CV study. Across the four versions of the payment card, large changes in the ranges (the highest values listed on the tested payment cards ranged from $200 to $10,000) and values listed in the center of the payment card (the center value in the test ranged from $16 to $125) did not substantively change the distribution of WTP responses. Further, a similar percentage of respondents selected the same WTP values regardless of which column in the payment card the WTP values were listed. This demonstrates that, contrary to the ARCO comments, the percent of respondents who answered values in each column of the payment card are happenstance findings of no consequence.[5]

Missing "No Response" Option

Dr. Hausman comments (page 43) that the payment card did not include a "no response" option, which he noted was initially suggested by the NOAA CV panel. This issue was discussed in detail on page 1-11 of Montana's CV report, where based on CV data in Chestnut and Rowe (1990) we conclude: "The omission of the 'no answer' response to the WTP question is expected to have an insignificant impact on the reliability of the results and to potentially downwardly bias the results." Both NOAA and DOI, when issuing their proposed CV regulations, significantly reduced the emphasis on this design issue as compared to the NOAA CV panel report. Dr. Hausman was aware of these facts but provides speculation rather than evidence to refute them.

---

[4] The exception to the result of no range bias was when the high end of the distribution of values respondents would desire to select is not included on the payment card. This results in a downward "truncation bias."

[5] Note also that the median values of the raw data in the Rowe, Schulze, and Breffle (forthcoming) study for the same payment card version as in the Montana study was $125 to clean up all hazardous waste sites in Colorado. In the Montana study the median was $40 to cleanup the upper Clark Fork sites, and the entire WTP distribution shows a similar downward shift of one column in the payment card. This also refutes the concept that respondents simply choose the center of the payment card.

6.3 INCENTIVE BIAS

Dr. Hausman (page 44) speculates that the $20 incentive fee may create a valuation anchor, which he obliquely infers may bias the stated WTP upward. This is a speculative comment for which he provides no supporting data or literature. The literature indicates that such "starting point" anchors may bias reported WTP from the true value toward the anchor. If that occurred in this study, then there is a downward bias because (net of $0 responses who are apparently not influenced by the $20 incentive) only 27 percent reported an initial WTP of less than $20, and 73 percent reported an initial WTP of more than $20.

He attempts to support his point by noting that 17 respondents wrote a comment indicating they would give their $20 to the cleanup, and that these individuals had a higher stated WTP than the remainder of the sample. Closer examination shows that his presentation misrepresents the data. Turning to the cleaned data used in the damage assessment, there are 15 individuals who "offered" to use their $20 to fund cleanup (9 who responded to questionnaire Version 1, and 6 who responded to questionnaire Version 2). The nine respondents of interest responding to Version 1 provided a higher average WTP than did the remainder of the Version 1 sample. But, compared to the rest of the Version 1 respondents, these nine respondents had a 33 percent higher average income, rated their responsibility to pay for clean-up higher than did the rest of the sample, and rated the importance of cleaning up the sites as more important than did the rest of the sample. All these factors account for their higher WTP responses. In Version 2, the six respondents of interest had nearly the same WTP and reported only slightly different income and ratings on cleanup, as compared to the remainder of the sample. The real significance of the written comments to return the $20 is that some respondents offered, without being asked, to put cash behind their responses.

CHAPTER 7
SAMPLING AND STATISTICAL ISSUES

Drs. Hausman and Desvousges provide a number of comments as to the statistical treatment of data in the State's CV study. The survey data and literature support the State's procedures.

7.1 NONRESPONSE BIAS CORRECTION

Nonresponse bias may arise if the nonrespondents to the CV survey hold significantly different values from the respondents, and if this difference is not addressed. The State's CV study corrects for potential nonresponse bias by comparing characteristics (age, income, and sex) for respondents to the mail survey and for a subset (88) of nonrespondents who were contacted by telephone and determining how these differences are predicted to influence WTP. Such procedures are standard practice in CV and survey research projects.

Drs. Hausman (pages 50-52) and Desvousges (pages 58-61) suggests that additional variables may also be relevant for analyzing nonresponse bias, and they speculate, without analysis, that the State's correction may not fully correct for potential nonresponse bias leading to upwardly biased damage estimates.

To place this issue in perspective, we can readily examine the impact of the most extreme assumptions about nonresponse bias. The first is that all nonrespondents have $0 WTP for cleanup of the upper Clark Fork sites. If this extreme assumption were used, the Version 1 mean WTP would be approximately 20 percent less than reported. The second extreme assumption is to assume that all nonrespondents have the same value as respondents, which would result in values 18 percent larger than reported by the State for both Versions 1 and 2. Neither of these extreme assumptions is supported by the data from the telephone survey for this study or by any other available data.

The State's documented nonresponse bias assessment, based on respondent data, results in a correction nearly halfway between the extreme assumptions. In the CV report Figures 5-1 and 5-2, the State demonstrates that the WTP values do not decline as the repeated survey contacts resulted in more response. Combined with the follow-up survey of nonrespondents, we conclude that even if there is residual nonresponse bias beyond that corrected for in the State's assessment, the residual bias would have a small effect on the population-wide damage estimates, and it is not clear whether any such residual bias would have an upward or downward effect on the damage assessment. In contrast, ARCO's experts provide speculation,

with no analysis, that the nonresponse bias correction may be insufficient and may bias the estimates upward.

7.2 THE STATE'S WTP REGRESSION ANALYSIS

The State uses a regression model that explains how WTP varies across individuals as a function of each individual's attitudes and opinions, motives, behavior, and socioeconomic characteristics. This model is used in the correction procedures undertaken to address sampling, nonresponse, and responsibility biases. ARCO expresses concerns with the statistical specification of the regression model. These comments are inaccurate and misleading and, overall, have little substantive effect on the damage assessment.

7.2.1 Tobit Model Specification

Because the data includes a substantial number of zero WTP observations, a Tobit regression specification was used by the State. Dr. Hausman notes (page 46) that the Tobit model underpredicts the number of zero observations in the data (the model predicts about 5 percent of the sample at $0 rather than the actual 13 percent of the sample). Dr. Hausman makes no claim about whether this is a significant issue or whether it biases upward or downward the analyses for which this model is used, nor does he provide an alternative analysis. In fact, it is misleading to suggest this finding has significance to the estimates of damage. As is well known, least squares regression analysis methods predict best around the mean values for the dependent variable (WTP) and the explanatory variables, and the predictions worsen as one moves away from the means. The fact that the model has errors in predicting the number of observations at the extreme tail of the distribution says little of substance about the overall fit of the model or its usefulness for predicting changes in the mean when correcting for sampling, nonresponse, and responsibility bias.

7.2.2 Indices

In the State's regression model, 14 different variables measuring respondents' attitudes and motives were aggregated into four indices[1] to streamline the model and to avoid multicollinearity problems. Dr. Hausman expresses concern with the use of the index

[1] An index SIS was the sum of the scores for the importance of cleaning up the four NPL sites in the upper Clark Fork Basin (Questions 19-22). An index IIS was the sum of the scores rating the importance of the issue of cleaning up hazardous waste sites (Questions 3, 7, and 11). An index ULS was the sum of the scores for the likely use of different upper Clark Fork resource groups (Questions 23-25). An index MS was the sum of the score for the four different motives (use, current bequest, future bequest, and existence) for cleaning up the sites.

variables and suggests the index variables should be replaced by the 14 individual variables (pages 46-47). He further asserts that when the 14 variables are included separately in the regression, their coefficients indicate that WTP is not correlated with respondents' rankings of importance for cleaning up the sites. From this he implies the WTP results are not consistent with our expectations and are therefore unreliable. The concerns raised are readily examined with the data and are found to be incorrect and seriously misleading.

First, the correction suggested by Dr. Hausman of including all 14 variables rather than the four indices has virtually no impact on the analyses for which the model is used (to correct for sampling, nonresponse, and responsibility bias). Further, the suggested alternative only slightly increases the statistical precision of the regressions (as is expected when replacing 4 variables with 14 variables) and raises predicted WTP about 6 percent. His revision has no impact on the estimated damages or on the reliability of the results.

Second, when all 14 variables in the four indices are included in the regression, the individual coefficients on these variables are generally not statistically significantly different from zero. This implies that the attitudes of respondents are not correlated with their stated WTP (a result that is in sharp contrast to the actual facts presented below). Dr. Hausman specifically notes that the "estimated coefficient on the Montana Pole site is -0.13, implying that the more important the respondent rated the Montana Pole site the lower his or her WTP answer." He neglects to note that this coefficient is not statistically different from zero (at even the lowest of statistical significance levels) and therefore does not imply what he asserts.

The fundamental problem with the Hausman argument is that he does not acknowledge multicollinearity. Multicollinearity is an issue when some variables in the regression are highly correlated. The result of multicollinearity is that it is difficult for the regression analysis to determine the separate impact of each individual highly correlated explanatory variable on the dependent variable (WTP). Multicollinearity often results in a regression with high explanatory power but with coefficients on the individual highly correlated variables that are unstable and often are not statistically significant, even when they are highly correlated with the dependent variable. This is the case here.

- The simple correlations between the individual variables in the site importance index (SIS including the importance of cleaning up Silver Bow Creek, the Montana Pole site, etc.) all exceed 0.8 and are highly statistically significant. The variables in the issue importance index (IIS) are all highly correlated at 0.68 or higher (and highly statistically significant), the variables in the use index are all highly correlated at 0.4 or higher (and highly statistically significant), and the variables of the motivation index (use by my household, use by others, future use, existence value) are all correlated at 0.49 or higher (and highly statistically significant). Further, the variables used in different indices are also highly correlated (generally from 0.3 to 0.8). When 14

variables all have such high simple correlations, one has strong multicollinearity.[2]

- The simple correlations between the site scores and WTP, for each of the four sites, are positive and statistically significant. This implies that as the site score increases, for each site, the WTP to clean up the upper Clark Fork sites also increases. The fact that the Hausman regression estimates insignificant, and even negative, coefficients for the individual site importance variables (implying that WTP should decrease with increases in ratings of the site) is again symptomatic of strong multicollinearity problems. Similarly, each of the individual resource cleanup importance variables and cleanup motivation variables are positively correlated with WTP, as would be expected.

Multicollinearity effectively invalidates tests of individual variables as predictors because the statistical test asks whether it is useful to add the particular variable as a predictor of WTP over and above all the other variables. If there are close substitute variables, as in the case of Dr. Hausman's analysis using 14 separate predictor variables, then the answer will almost surely be that it is not useful to add the particular variable. This does not imply that the particular variable is not related to WTP; rather, it simply implies in this case that there are substitute variables in the model. This multicollinearity indicates that respondents responded in similar patterns to similar questions, which again indicates that respondents were answering the questions carefully and consistently.

Standard practice is to combine substitute variables into indices (as done by the State). The resulting indices will have less multicollinearity thereby facilitating easier interpretation and more reliable statistical estimates. Furthermore, multicollinearity of the variables within an index has the effect of making the index score significantly more reliable and stable than any of the individual variables. Dr. Hausman has completely misrepresented the data by clearly inappropriate data analysis that simply ignores, rather than addresses, the strong multicollinearity in the data.

7.3 DATA TREATMENT ISSUES

Dr. Desvousges raises several issues with the treatment of data in the State's CV study. These concerns are not substantive.

[2] Given that virtually all of the simple correlations among the 14 variables exceed the regression R^2, multicollinearity is particularly strong.

7.3.1 Missing Data

First, Dr. Desvousges comments that 5 percent of respondents had missing responses to the second WTP question (Question 33), leading to increased uncertainty in the damage estimates (page 53). Dr. Desvousges misstates the data because only 1 percent of respondents who had nonzero WTP had missing values for Question 33 (the other 4 percent had $0 WTP, and so the Q33 response is immaterial as anything times $0 is still $0). Either way, because of the small number of missing observations, the selection of a method to deal with missing observations has *de minimis* impact on the overall results. Rather than delete all the information for the few individuals with missing observations for one or a few variables, the state employed well-accepted standard statistical procedures. For this specific variable, a "zero-order missing value replacement" procedure of substituting the sample mean value for the missing value was used. Because of the very limited number of observations involved (1 percent), this standard procedure adds little error to the damage assessment. Dr. Desvousges's attempt to raise concerns about error in this procedure is unwarranted and misleading because this is a trivial issue.

7.3.2 Coding of Second WTP Responses

Dr. Desvousges raises a concern that the State altered the second WTP for 77 respondents and says the State's procedure "is not based on any empirical evidence" (pages 53-54). The State's treatment of the data on this issue has minimal impact on the results, and the survey data support the procedures employed by the State.

The specific issue is that 77 respondents (only 67 in cleaned data) provide WTP responses to Questions 28 and 31 that imply they value complete cleanup less than partial cleanup. In Version 1, where WTP for complete cleanup was asked first, four respondents (1 percent) listed that their WTP for partial cleanup was "more than 100%" of their WTP for complete cleanup. In Version 2, where WTP for partial cleanup was asked first, 73 respondents (63 in the cleaned data) answered that their WTP for complete cleanup was "less than 100%" of their WTP for partial cleanup. In each case, the WTP results for these individuals suggest that partial cleanup is valued higher than complete cleanup. In order to calculate a specific value for these individuals, the State set the partial and complete WTP as equal. The reasons the State did this were provided to ARCO in deposition but ignored in their response. There is empirical evidence and good reasoning to support replacing "more than 100%" and "less than 100%" responses with a value of 100%. For these individuals the mean rankings of their satisfaction of a complete cleanup are statistically significantly higher than for partial cleanup for each of the resource groups: groundwater, surface water, and soils. As a result, they indicate they are more satisfied with complete cleanup than partial cleanup. Equaling the WTP for partial and complete cleanup is a conservative procedure.

The reason some individuals respond in a manner that suggests partial cleanup may be valued more than complete cleanup is intuitively appealing. They expect complete cleanup will not work and/or will be unreasonably expensive and therefore should not be supported. This cost-benefit reasoning is reflected in the comments of respondents in the verbal protocols, pretests, and final surveys (see Appendix 8A).

Finally, it is important to note that these corrections have a very small impact on the overall results. The 67 respondents of interest provided a first WTP response that was less than half as large as the remaining 90 percent of the sample. Even doubling or halving the second imputed WTP responses for these individuals would change the overall results for the second WTP by less than 3 percent.

7.3.3 Protest Zeros and Large WTP Responses

Dr. Desvousges raises a generic concern about the removal of protest zeros and suggests that fewer $0 WTP responses might have been removed with alternative criteria (pages 56-57). In fact, the State's criteria are more stringent than is typically the case in the literature (for example, our procedure is similar to but more conservative than in recent CV work by Desvousges (as reported in Boyle et al., 1994) and we probably err toward including invalid zeros and reducing estimated WTP (as explained in detail on pages 5-2 through 5-4 of the State's report). Even in the *most extreme*, and incorrect, case of retaining all reported $0s as valid $0 WTP, the sample wide WTP mean value would be about 7 percent less than reported.

Dr. Desvousges raises concerns about the procedures to identify and remove potentially suspect large WTP responses (pages 57-58). In fact, the state used a procedure that parallels procedures used by Dr. Desvousges in his own work (Desvousges et al., 1987; Boyle et al., 1994), and we use a statistical cutoff that is biased toward removing potentially valid large WTP responses. Dr. Desvousges makes a comment that the State's experts did not use a simple procedure used in their earlier studies, but he fails to note that the procedure used in the State's work is based on a more comprehensive evaluation and is more conservative toward eliminating large WTP responses than in recent CV work by Desvousges (Boyle et al., 1994).[3]

[3] In some earlier studies, the State's experts had removed responses where the WTP exceeded 1 percent of income. In this study some individuals remain in the sample with WTP in excess of 1 percent of income. This often occurs because these individuals express high levels of importance for cleanup of the site, provide high rankings on expected use of site resources, and have other characteristics consistent with high values. Many individuals specifically write in comments supporting and explaining the listed WTP amount. Finally, after adjusting for embedding, in Version 1 the largest WTP for cleanup of the Clark Fork sites is $240 (three respondents), and in Version 2 the largest WTP for cleanup of the Clark Fork sites is $800 (one respondent). These respondents generally live near the site and provide high importance

7.4 RESPONSIBILITY BIAS CORRECTION

CV survey respondents may partially or wholly reject the CV scenario if, for example, they feel the polluter should pay. As a result, respondents may report a reduced WTP that only partially reveals their total value of cleanup. This introduces a downward "responsibility bias" into the analysis. The State conducts a statistical analysis to measure and correct for "responsibility bias" (see pages 5-34 through 5-36 of the State's CV report).

Dr. Hausman asserts that the State's statistical specification is incorrect because of potential sample selection bias. He then presents an alternative statistical analysis and concludes there should be no responsibility bias correction (pages 49-50, Dr. Desvousges provides comments on this issue on pages 58-60 but performs no analysis).

Dr. Hausman's sample selection model has significant econometric limitations and is misleading in its intent to show that responsibility bias should not affect the State's calculations. This applies to both the traditional and semiparametric approaches he identifies. Dr. Hausman's selection model is based on the observation that sample selection can be viewed as a missing variable problem (Heckman, 1979). The procedure has two steps. In the first step, a probit or logit equation is fit to a set of exogenous explanatory variables, at least one of which cannot logically be included in the WTP equation for the purpose of identification. Estimated coefficients from this model are used to predict the variable missing from the WTP equation (the "inverse Mills ratio variable"). This predicted variable is added to the WTP regression in the second step. He then focuses on the variable RESPONSIBILITY in models with (Hausman model) and without (State model) his selection correction. The estimated coefficient on RESPONSIBILITY is statistically significant in the State's model, and based on this the State makes a responsibility bias correction. But the estimated coefficient on RESPONSIBILITY is statistically insignificant in the Hausman selection model, which he uses to infer that no responsibility bias should be made. There are three problems with Dr. Hausman's conclusions from this analysis:

1. As an exercise in correcting for sample selection bias, his conclusion should have been that there is *no* sample selection bias present, based upon the low t-statistic (not reported in his Table 5) on the added variable. This t-statistic is barely statistically significant at a level that Dr. Hausman elsewhere suggests is only "occasionally used" (page 33).

2. His selection methodology is flawed because his only identifying variable in the first step, OTHERPAY, is *not* exogenous. Therefore, his two-equation selection

ratings for cleanup of the site and all sites in general. See also the State's CV report Table 5-20, which further illustrates the validity of retained large WTP responses.

model is not identified, which means there is no interpretation possible for the coefficient of RESPONSIBILITY (or any other variable).

3. His chosen identifying variable, OTHERPAY, is not only endogenous, but is also conveniently highly positively correlated with RESPONSIBILITY. By using OTHERPAY in forming the predicted missing variable, Hausman is guaranteeing that the coefficient on RESPONSIBILITY will be smaller (the multicollinearity problem discussed in Section 7.2.2).

We could not evaluate Dr. Hausman's semiparametric tests used in this section of his report because supporting materials (including the computations themselves, which are important because nothing in this forthcoming paper is related to the sample selection bias issue being addressed in Dr. Hausman's comments) were not provided to the State of Montana. However, experience indicates that these tests are based on the same types of procedures as discussed above, and that the statistical power of such tests are often very low. As the authors of the Hausman and Newey paper note (page 1) "often nonparametric models" may be difficult to interpret, and their usefulness in applied research has been demonstrated in a limited number of cases" (parenthetical phrase added for clarity).

APPENDIX 7A
SAMPLE PRETEST COMMENTS CONCERNING PARTIAL VERSUS COMPLETE CLEANUP

"From what I can see so far, I would suggest that you do just a partial cleanup as, the price tag for this state would be staggering!!" Raw partial = Raw complete WTP = $40. 11/05/92-V1, S1, 003.

"It needs to be stressed that complete restoration is unattainable. There isn't enough money in the world to return the Clark Fork into a Missouri-type fishery. There are past transgressions we will have to live with until the next glaciers scour it away." Complete=partial WTP = $650. 11/05/92-V2, S2, 004.

100% for partial cleanup, "Only if complete cleanup is absolutely impossible!"; $450 WTP; 12/06/92 - V1; 6

"...In my own mind partial cleanup in many cases would be sufficient..."; Satisfaction of complete cleanup and partial cleanup of surface water were both given 5's; 100% for partial cleanup; $10 WTP;12/06/92 - V1; 3

Satisfaction of partial cleanup of surface water given 6 while complete received a 5; Both partial and complete satisfaction for groundwater given 3's; "DON'T KNOW" WTP; 12/06/92 - V1; 2

"...Total restoration might not be the best bang for the buck..."; "Stabilization of all of the sites is 1st priority (incl. Containment)"; $200 WTP; 12/15/92 - V2; 6

"For surface water satisfaction, gave complete a 6 and partial a 6 as well; WTP 100% of partial for total (same); $25 WTP; 12/15/92 - V2; 2

"If they tried and they still couldn't do all of it, but they were willing to try, then I'd be willing to pay the same amount. 100%" Partial WTP = Complete WTP. V.P. 2

"I don't think that it is necessary to restore the area completely to a pre mining condition. We would be willing to pay $450 for a realistic approach such as outlined in question #20"; responded 100% for Q20 (partial rated equally to complete); $200 WTP for complete *willing to pay more for partial; 12/14/92 - V3; 4

"I would be willing to pay $120/year [for partial]. The way this survey is structured needs to be changed. If you answer $0 in Q11 then your answer automatically has to be zero in Q20. We can't afford the cost of a complete cleanup."; $0 WTP for complete; 12/14/92 - V3; 6

CHAPTER 8
OTHER CONTINGENT VALUATION ISSUES

Other miscellaneous comments about the development and use of the contingent valuation analysis were raised in the ARCO comments on the CV analysis and on the use of the State's CV study in the Damage Determination report. These other comments are addressed here.

8.1 CONSERVATIVE ESTIMATES

Dr. Hausman speculates that the State's CV results may not be conservative, even though the State's CV results are based on conservative procedures (pages 52-53). He bases his arguments on a "suppose the initial average WTP answer was 100 times greater than true economic value" comment and discussion of the effect of just one of the State's conservative procedures, which in itself reduces values by 17 percent. Under this scenario, he suggests that values may still be 83 times too large.

This presentation lacks any foundation in terms of the CV study results. Dr. Hausman speculates, without foundation, that the CV results may be 100 times too large, or equivalently that the average Montana household values for cleaning up the upper Clark Fork sites should be $0.47 per year for 10 years (or a present value of under $4, which is apparently consistent with his understanding and valuation of the injuries in Montana). This absurd assumption is totally inconsistent with the stated opinions, values, and behaviors of Montana residents (see Section 3.0). Further, he focuses on just one of the many conservative procedures used by the State throughout the derivation of the CV results. It is difficult to imagine a more extreme, and invalid, presentation being made.

8.2 AGGREGATE DAMAGE COMPUTATIONS

In Section VIII of Dr. Hausman's report, he makes comments on the aggregate damage determination, briefly addressed in Chapter 6 on the State's CV report and addressed in detail in the State's Damage Determination report. The response to these comments is provided in Appendix B of the State's revised Damage Determination report.

CHAPTER 9
REFERENCES

Adamowicz, V.L. and T. Graham-Tomasi. 1991. "Revealed Preference Tests of Non Market Goods Valuation Methods". *Journal of Environmental Economics and Management.* **20**(1):29-45.

Arrow, K., E. Leamer, H. Schuman, and R. Solow. 1994. "Subject Entitled: NOAA Proposed Rule on Natural Resource Damage Assessments. Memorandum to NOAA Damage Assessment Regulation Team in Response to ANPRM 1-7-94. Comment Number 69." October.

Arrow, K., R. Solow. E. Leamer, R. Radner, and H. Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation," *Federal Register.* **58**(10):4602-4614.

Birnhaum, M.H. 1982. "Controversies in Psychological Measurement". In *Social Attitudes and Psychophysical Measurement.* B. Wegener (ed.). Lawrence Erlbaum Associates, Hillsdale, NJ.

Bishop, R.C. and T.A. Heberlein. 1979. "Measuring Values of Extra-Market Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics.* **61**(5):926-930.

Bishop, R.C. and T.A. Heberlein. 1990. "The Contingent Valuation Method." In *Economic Valuation of Natural Resources: Issues, Theory and Applications.* Johnson & Johnson (eds). Westview Press: Boulder, CO. pp. 81-104.

Blomquist, G.C. 1988. "Valuing Urban Lakeview Amenities Using Implicit and Contingent Markets". *Urban Studies.* **25**:333-340.

Boyce, R., G.H. McClelland, T. Brown, G. Peterson, and W. Schulze. 1992. "An Experimental Examination of Intrinsic Values as a Source of the WTA-WTP Disparity"." *American Economic Review.* **82**(5):1366-1373.

Boyle, K.J. and R.C. Bishop. 1988. "Welfare Measurements Using Contingent Valuation: A Comparison of Techniques". *American Journal of Agricultural Economics.* **70**(1):20-28.

Boyle, K.J., W.H. Desvousges, F.R. Johnson, R.W. Dunford, and S.P. Hudson. 1994. "An Investigation of Part-Whole Biases in Contingent-Valuation Studies". *Journal of Environmental Economics and Management.* **27**(1):64-83.

Brookshire, D.S. and D.L. Coursey. 1987. "Measuring the Value of A Public Good: An Empirical Comparison of Elicitation Procedures." *American Economic Review*. 77:554-566.

Brookshire, D.S., M.A. Thayer, W.D. Schulze, and R.C. d'Arge. 1982. "Valuing Public Goods: A Comparison of Survey and Hedonic Approaches". *American Economic Review*. 72:165-177.

Brookshire, D.S., M.A. Thayer, J. Tschirhart, and W.D. Schulze. 1985. "A Test of the Expected Utility Model: Evidence from Earthquake Risks". *Journal of Political Economy*. 93:369-389.

Browner, C. 1994. Letter prepared by Carol Browner, Administrator, U.S. Environmental Protection Agency, Washington, D.C., March 21, 1994.

Carson R.T., N.E. Flores, K.M. Martin, and J.L. Wright. 1994. "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods." University of California, San Diego, Department of Economics Discussion Paper 94-07.

Carson, R.T., R. Mitchell, W. Hanemann, R. Kopp, S. Presser, and P. Ruud. 1992. "A Contingent Valuation Study of Lost Passive Use Values Resulting from the Exxon Valdez Oil Spill". A Report to the Attorney General of the State of Alaska.

Chestnut, L.G. and R.D. Rowe. 1990. "Preservation Values for Visibility Protection in the National Parks". Prepared by RCG/Hagler Bailly for the U.S. EPA, Office of Air Quality Planning and Standards, NC, and the Air Quality Management Division, National Park Service, Denver. U.S. EPA Agreement No. CR813686 with the University of Colorado. February.

Coursey, D.L., J.J. Hovis, and W.D. Schulze. 1987. "The Disparity between Willingness to Pay Measures of Value". *Quarterly Journal of Economics*. August. pp. 680-690.

Cummings, R.G., D.S. Brookshire, and W.D. Schulze. 1986. *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*. Rowman & Littlefield: Savage, MD. 270 p.

d'Arge, R. and J.F. Shogren. 1989. "Non-Market Asset Prices: A Comparison of Three Valuation Approaches". In *Valuation Methods and Policy Making in Environmental Economics*. H. Folmer and E. van Ireland (eds.). Elsevier Publishers: Amsterdam.

Davis, D.D. and C.A. Holt. 1993. *Experimental Economics*. Princeton University Press: Princeton, NJ.

Desvousges, W.H. 1995. Volume 1: Critique of the State of Montana's Contingent Valuation Study. Expert report submitted to United States District Court, District of Montana, Helena Division, in regard to State of Montana v. Atlantic Richfield Company. No. CV-83-317-HLN-PGH. July 13.

Desvousges, W., V.K. Smith, and A. Fisher. 1987. Option Price Estimates for Water Quality Improvements: A Contingent Valuation Study for the Monongahela River. *Journal of Environmental Economics and Management.* **14**(3):148-267.

Desvousges, W.H., V.K. Smith, and M.P. McGivney. 1983. "A Comparison of Alternative Approaches for Estimating Recreation and Related Benefits of Water Quality Improvements." Report to the U.S. EPA by Research Triangle Institute.

Desvousges, W.H., F.R. Johnson, R.W. Dunford, K.J. Boyle, S.P. Hudson, and K.N. Wilson. 1992. "Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability." Paper presented at the symposium "Contingent Valuation, A Critical Assessment," Washington, DC, April 2-3.

Diamond, P., J.A. Hausman, G. Leonard, and M. Denning. 1992. "Does Contingent Valuation Measure Preferences: Experimental Evidence?" Paper presented at the symposium "Contingent Valuation, A Critical Assessment," Washington, DC, April 2-3.

Dickie, M., A. Fisher, and S. Gerking. 1987. "Market Transactions and Hypothetical Demand Data: A Comparative Study." *Journal of the American Statistical Society.* **82**:69-75.

Dolan, E.G. 1977. *Basic Economics.* The Dryden Press: Hinsdale, IL.

Duffield, J.W. and D.A. Patterson. 1991. "Field Testing Existence Values: An Instream Flow Trust Fund for Montana Rivers," Paper presented at the Allied Social Science Associations Meeting, New Orleans, January.

Duffield, J.W. and D.A. Patterson. 1992. Field Testing Existence Values: An Instream Flow Trust Fund for Montana Rivers, mimeo, University of Montana.

Green, D. K. Jacowitz, D. Kahneman, and D. McFadden. 1995. Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods, mimeo, University of California, Berkeley.

Hagel, J.H. and A.E. Roth. 1995. *The Handbook of Experimental Economics.* Princeton University Press: Princeton, NJ.

Hanemann, W.M. 1994. "Valuing the Environment through Contingent Valuation." *Journal of Economic Perspectives.* **8**(4):19-43.

Hausman, J.A. 1995. Expert report submitted to United States District Court, District of Montana, Helena Division, in regard to State of Montana v. Atlantic Richfield Company. No. CV-83-317-HLN-PGH. July 13.

Hausman, J.A. and P. Diamond. 1993. "On Contingent Valuation Measurement of Nonuse Values." In *Contingent Valuation: A Critical Appraisal.* J.A. Hausman (ed.).

Hausman, J.A. and P. Diamond. 1994. "Contingent Valuation: Is Some Number Better than No Number?" *Journal of Economic Perspectives.* **8**(4).

Hausman, J.A. and W.K. Newey. 1995. "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss". MIT, Department of Economics, Working Paper, January.

Hausman J.A., P. Diamond, G. Leonard, and M. Denning. 1993. "Does Contingent Valuation Measure Preferences?" In *Contingent Valuation: A Critical Appraisal.* J.A. Hausman (ed.).

Heckman, J. 1979. "Sample Selection Bias as a Specification Error". *Econometrica.* **47**:153-161.

Irwin, J.R., G.H. McClelland, and W.D. Schulze. 1992. "Hypothetical and Real Consequences in Experimental Auctions for Insurance against Low Probability Risks". *Journal of Behavioral Decision Making.* 5:107-116.

Isaac, R.M., K. McCue, and C. Plott. 1985. "Public Goods Provision in an Experimental Environment". *Public Choice.* **62**(3):217-236.

Johnson, B. 1993. Memorandum re: Comments on SAB Environmental and Economics Advisory Committee (EEAC) Review of Agency's Contingent Valuation Study of Groundwater". Prepared by Barnes Johnson, Chief, Regulatory Analysis Branch, U.S. EPA, Office of Solid Waste and Emergency Response, Washington D.C., October 20, 1993.

Johnson, R.L., N.S. Bregenzer, and B. Shelby. 1990. "Contingent Valuation Question Formats: Dichotomous Choice versus Open-Ended Responses". In *Economic Valuation of Natural Resources: Issues, Theory, and Applications.* Johnson, R.L. and G.V. Johnson (eds.). Westview Press: Boulder, CO.

Kealy, M.J. and R.W. Turner. 1993. "A Test of the Equality of Closed-Ended and Open-Ended Contingent Valuations." *American Journal of Agricultural Economics.* **75**:321-331.

Kim, O. and M. Walker. 1984. "The Free-Rider Problem: Experimental Evidence". *Public Choice.* **43**:3-24.

Knetsch, J.L. and R.K. Davis. 1966. "Comparison of Methods for Recreation Evaluation." In *Water Research*. A.V. Kneese and S.C. Smith (eds.). Johns Hopkins Press: Baltimore.

Loomis, J.B., D.F. Sorg, and D.M. Donnelly. 1986. "Evaluating Regional Demand Models for Estimating Recreation Use and Economic Benefits: A Case Study". *Water Resources Research*. **22**:431-438.

Maler, K.G. 1974. *Environmental Economics*. Johns Hopkins Press: Baltimore, MD.

McClelland, G.H., W.D. Schulze, and D.L. Coursey. 1993. "Insurance for Low-Probability Hazards: A Bimodal Response to Unlikely Events". *Journal of Risk and Uncertainty*. 7:95-116.

McConnell. 1993. "Indirect Methods for Assessing Natural Resource Damages under CERCLA." *Valuing Natural Assets: The Economics of Natural Resource Damage Assessments*. R.J. Kopp and V.K. Smith (eds.) Resources for the Future. Washington, D.C.

McFadden, D. 1994. "Contingent Valuation and Social Choice," *American Journal of Agricultural Economics*. **76**(4):689-708.

Mendelsohn, R., D. Hellerstein, M. Huguenin, R. Unsworth, and R. Brazee. 1992. "Measuring Hazardous Waste Damages with Panel Models." *Journal of Environmental Economics and Management*. **22**(3):259-271.

Mitchell, R.C. and R.T. Carson. 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Resources for the Future, Washington, DC. 463 p.

Morey, E.R., R.D. Rowe, D. Waldman. Revised Report and Rebuttal: Assessment of Damages to Anglers and Other Recreators from Injuries to the Upper Clark Fork River Basin. Prepared for the State of Montana Natural Resource Damage Litigation Program by Hagler Bailly Consulting, Inc. October.

Nickerson, C.A. and G.H. McClelland. 1989. "Across-Persons vs. Within-Persons Tests of Expectancy-Value Methods: A Methodological Note. *Journal of Behavioral Decision Making*. Vol. 2:261-270.

Rowe, R.D., W.D. Schulze, and W.S. Breffle. "A Test for Payment Card Biases." *Journal of Environmental Economics and Management*. Forthcoming, May 1996.

Rowe, R.D. W.D. Shaw, and W.D. Schulze. 1992. Nestucca Oil Spill. In *Natural Resource Damages: Law and Economics*. K.M. Ward and J.W. Duffield (eds.). John Wiley and Sons, Inc.: New York City, NY. pp. 527-554.

Rowe, R.D., W.D. Schulze, D. Schenk, and L.G. Chestnut. 1991. Contingent Valuation of Natural Resource Damage Due to the Nestucca Oil Spill. Prepared by RCG/Hagler Bailly for the State of Washington, Department of Wildlife (Olympia), British Columbia Ministry of Environment (Victoria), and Environment Canada (Vancouver). June.

Schulze, W., G. McClelland, and B. Hurd. 1986. *A Case Study of a Hazardous Waste Site: Perspectives from Economics and Psychology. Draft Report.* Prepared for the U.S. Environmental Protection Agency, Washington, D.C. Contract #CR812504-01-1.

Schulze, W.D., G. McClelland, D. Waldman, and J. Lazo. 1995. "Sources of Bias in Contingent Valuation". In: *The Contingent Valuation of Environmental Resources: Methodological Issues and Research Needs.* D. Bjornstad and J. Kahn, Chapter 6. (eds.). Edward Elgar Publishing Ltd.: Cheltenham, U.K. Forthcoming.

Schulze, W.D., R.D. Rowe, W.S. Breffle, R.R. Boyce, and G.H. McClelland. 1995. Contingent Valuation of Natural Resource Damages due to Injuries to the Upper Clark Fork River Basin. Prepared for the State of Montana Natural Resource Damage Litigation Program. January.

Seip, K. and J. Strand. 1992. "Willingness to Pay for Environmental Goods in Norway: A Contingent Valuation Study with Real Payment". *Environmental and Resource Economics.* **2**: 91-106.

Seller, C., J.R. Stoll, and J.P. Chavas. 1985. "Validation of Empirical Measures of Welfare Change: A Comparison of Nonmarket Techniques". *Land Economics.* **61**(2):156-175.

Shogren, J.F., S.Y. Shin, D.J. Hayes, and J.B. Kliebenstein. 1994. "Resolving Differences in Willingness to Pay and Willingness to Accept". *American Economic Review.* **84**:255-270.

Smith, V.K. 1993 "Nonmarket Valuation of Environmental Resources: An Interpretative Appraisal." *Land Economics.* **69**(1):1-26.

Smith, V.K. and W.H. Desvousges. 1986. "The Value of Avoiding a Lulu: Hazardous Waste Disposal Sites." *Review of Economics and Statistics.* **68**(2):293-299.

Thayer, M.A. 1981. "Contingent Valuation Techniques for Assessing Environmental Impacts: Further Evidence." *Journal of Environmental Economics and Management.* **8**:27-44.

Thayer, M., H. Albers, and M. Rahmatian. 1993. "The Benefits of Reducing Exposure to Waste Disposal Sites: A Hedonic Housing Value Approach." *Journal of Real Estate Research.*

Tietenberg, T.H. 1994. *Environmental Economics and Policy.* N.Y. Harper Collins.

Tversky, A. and D. Kahneman. 1974. "Judgment Under Uncertainty: Heuristics and Biases". *Science*. **185**:1124-1131.

U.S. Environmental Protection Agency (U.S. EPA). 1994. Comments on Proposed NOAA/DOI Regulations on Natural Resource Damage Assessment. Office of Policy, Planning, and Evaluation, Washington, DC. October 7.

Walsh, R.G., D.M. Johnson, and J.R. McKean. 1992. "Benefit Transfer of Outdoor Recreation Demand Studies, 1968-1988," *Water Resources Research* **28**(3):707-713.

Walsh, R.G., L.D. Sanders, and J.R. McKean. 1990. "The Consumptive Value of Travel Time on Recreation Trips." *Journal of Travel Research*. **29**:17-24.